IMPROVING GENE THERAPY ANALYSIS WITH MULTI-WAVELENGTH ANALYTICAL ULTRACENTRIFUGATION METHODS

AMY HENRICKSON
Bachelor of Science, University of Lethbridge, 2017

A thesis submitted
in partial fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

BIOMOLECULAR SCIENCE

Department of Chemistry and Biochemistry
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

IMPROVING GENE THERAPY ANALYSIS WITH MULTI-WAVELENGTH ANALYTICAL ULTRACENTRIGUATION METHODS

Amy Henrickson

Date of Defense: September 5th, 2023

Dr. Borries Demeler Thesis Supervisor	Professor	Ph.D.
Dr. Trushar Patel Thesis Examination Committee Member	Associate Professor	Ph.D.
Dr. Marc Roussel Thesis Examination Committee Member	Professor	Ph.D.
Dr. Jean-Denys Hamel Internal External Examiner Department of Chemistry	Assistant Professor	Ph.D.
Dr. Stephen Lodmell External Examiner University of Montana	Professor	Ph.D.
Dr. Michael Gerken Chair, Thesis Examination Committee	Professor	Ph.D.

Abstract

Over the past few years, research into gene therapies has dramatically increased, with thousands of drug candidates in clinical trials. However, only a few are available on the market today, highlighting the need for improved analysis methods that can help validate the drug development process. The challenges associated with gene therapy analysis vary with different vector types and compositions. The main challenges for lipid nanoparticles (LNPs), a non-viral vector, are their large size and inherent heterogeneity. The challenge for adeno-associated viruses (AAVs), a viral vector, is differentiating between the full AAV capsid and the product-related impurities due to their similar hydrodynamic radii and surface properties. For both vectors, these challenges prevent their accurate quantification and characterization by many chromatography and size-based techniques. We have employed analytical ultracentrifugation (AUC) to improve gene therapy analysis by characterizing solutes in a sample based on their molar mass, shape, and density. AUC provides high statistical certainty through bulk observations, resulting in the correct assessment of sample purity and loading states. To further improve the resolution of AUC results, we have incorporated multi-wavelength capabilities into our AUC methods, adding an orthogonal optical characterization dimension. This thesis presents considerations for the design, execution, and analysis of multi-wavelength (MW) AUC experiments, looking at cases where optical deconvolution is not possible and cases where it is. It also includes a section on using MW sedimentation velocity experiments to characterize protein-DNA interactions, providing an example of how stoichiometry can be determined from MW-AUC experiments. Further, we apply MW capabilities to several AUC methods to improve the quantification and characterization of AAVs and LNPs. This results in the precise quantification and characterization of vectors, product-related impurities, and other contaminants.

Contributions of Authors

This thesis is composed of multiple manuscripts which resulted from multiple collaborations.

Chapter 2, entitled "Multi-wavelength analytical ultracentrifugation of biopolymer mixtures and interactions," was published in *Analytical Biochemistry,* Volume 652, article 114728, on May 26, 2022. I performed all AUC experiments and analyses and assisted with writing the manuscript with Dr. Borries Demeler, who also conceived the experiments. Gary E. Gorbet, Dr. Alexey Savelyev, and Minji Kim contributed to the multi-wavelength analysis software development, Dr. Jason Hargreaves provided the oil seed protein samples, and Sarah K. Schultz and Dr. Ute Kothe prepared and contributed the fluorescent protein samples and edited the manuscript.

Chapter 3, entitled "Protein – DNA interactions characterized by multi-wavelength AUC," was published in *Nature Communications*, volume 13, article number 7724 (2022). The work was performed in collaboration with Ishtiyaq Ahmed, Dr. Jeanette Hahn, Dr. Faisal Tarique Khaja, Dr. David Dubnau, and Dr. Matthew B. Neiditch from Rutgers University, who developed the project and performed x-ray crystallography, EMSA, and *in vivo* experiments. Dr. Demeler and I conceived the AUC experiments and assisted with writing the AUC sections. I performed and analyzed the AUC data. In this chapter, I have copied the manuscript sections I helped write, with permission from the corresponding author, and rewrote the introduction and discussion sections.

Chapter 4, entitled "Density matching multi-wavelength analytical ultracentrifugation to measure drug loading of lipid nanoparticle formulations," was published in *ACS Nano,* Volume 15(3), pp. 5068-5076, on February 22, 2021. Dr. Jayesh Kulkarni and I share first authorship. I performed and analyzed all AUC experiments and assisted with the experimental design and writing of the manuscript with Dr. Borries Demeler, who also conceived the project. Dr. Jayesh Kulkarni, Dr. Josh Zaifman, and Dr. Pieter Cullis prepared and contributed the LNPs and performed TEM and DLS measurements, and Dr. Gary Gorbet developed the density matching software in UltraScan.

Chapter 5, entitled “Evaluation of multi-wavelength analytical ultracentrifugation for the characterization and quantification of AAV capsid loading states with UltraScan” was accepted by *Nanomedicine* on September 5, 2023, prepared in collaboration with Dr. Xiaozhe Ding, Austin G. Seal, Dr. Zhe Qu, Dr. John Forsey, Dr. Viviana Gradinaru, and Dr. Kazuhiro Oka who all contributed AAVs. Dr. Xiaozhe Ding also performed TEM at Caltech and Dr. Lauren Tomlinson performed the mass photometry experiments. I performed and analyzed the AUC data, and with the assistance of Dr. Borries Demeler, we conceptualized the experiments and wrote the manuscript.

Acknowledgments

It seems fitting that my first acknowledgment would be to Dr. Ute Kothe for introducing me to Dr. Borries Demeler; without both of you, I may never have started a Ph.D. Thank you, Dr. Demeler, for all your support and guidance throughout my work. I started with Dr. Demeler as a lab technician in September 2018, when he moved his lab from Texas to Lethbridge. I had only heard about analytical ultracentrifugation one week before meeting Bo. Still, I managed to convince him that I was an excellent fit for his lab. Two years later, he convinced me I would be an excellent Ph.D. student, both of which turned out to be great opportunities, so thanks for taking a chance on me and supporting and teaching me throughout the years. I am thankful to my past and present committee members, Dr. Trushar Patel, Dr. Sandy Ross, and Dr. Marc Roussel, for their advice and support and for helping me achieve this goal. Next, I must thank all of the collaborators I have had the privilege of working with throughout my time in Dr. Demeler's lab. You have all helped expand my knowledge and make me a well-rounded researcher.

To my family for always supporting and motivating me through whatever I chose to do, whether it be backpacking through foreign countries for months immediately before my sister's wedding (thanks for supporting that Ash), pursuing a Ph.D., and even moving to the USA. You have taught me the value of working and playing hard and that one of the best places to play is by a body of water with a drink in your hand, even if it is a pool made of straw bales.

Finally, to my fiancé, Tyler, you have seen me through some hard times and convinced me not to drop out a couple of times. Thank you for always being there for me, supporting me, and believing in me. I do not know where I would be without you, but I am very excited to be in Boston with you soon, and I cannot wait for our life together. I also cannot leave out our dog Ellie, who is ready to celebrate any accomplishment, even if champagne popping scares her.

Table of Contents

List of Tables

List of Figures

Supplemental Figures

List of Abbreviations

2DSA	Two-Dimensional Spectrum Analysis
2DSA-IT	Iterative Two-Dimensional Spectrum Analysis
AAV	Adeno-Associated Virus
ABDE	Analytical Buoyant Density Equilibrium
AEC	Anion Exchange Chromatography
AF4	Asymmetric Field Flow Fractionation
ASTFEM	Adaptive Space-Time Finite Element Method
AUC	Analytical Ultracentrifugation
BSA	Bovine Serum Albumin
CCD	Charge Coupled Device
ComEA$_{Bs}$	Comea From *B. subtilis*
ComEA$_{Gs}$	Comea From *G. stearothermophilus*
DLS	Dynamic Light Scattering
DSPC	Distearoylphosphatidylcholine
EMSA	Electrophoretic Mobility Shift Assay
FDA	Food And Drug Administration
FRET	Fluorescence Resonance Energy Transfer
GMP	Good Manufacturing Processes
HDX	Hydrogen-Deuterium Exchange
HEK 293	Human Embryonic Kidney 293 Cells
ISSF	Initial Simulated Scan Files
ITR	Inverted Terminal Repeats
LIMS	Laboratory Information Management System
LNP	Lipid Nanoparticle
MP	Mass Photometry
mRNA	Messenger RNA
MW	Multiwavelength
MW-AUC	Multiwavelength Analytical Ultracentrifugation
NNLS	Non-Negative Least Squares
OD	Optical Density
OD	Oligomerization Domain
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PCSA	Parametrically Constrained Spectrum Analysis
PEG	Polyethylene Glycol
PSV	Partial Specific Volume
Rh	Hydrodynamic Radius
RMSD	Root Mean Square Deviation
SEC	Size Exclusion Chromatography
siRNA	Short Interfering RNA

SV	Sedimentation Velocity
SVE	Sedimentation Velocity Experiment
TEM	Transmission Electron Microscopy
TFP	mTeal Fluorescent Protein
UFP	Ultramarine Fluorescent Protein
UV	Ultraviolet
VP	Viral Protein

Chapter 1. Introduction

1.1 Analytical Ultracentrifugation

AUC is a first principle technique that can provide hydrodynamic and thermodynamic information on samples. AUC has played an essential role throughout the years, especially in characterizing proteins and protein complexes, and providing information on sample purity, macromolecular partial concentrations, oligomerization, aggregation, and stability [1-4]. Although this field has had its ups and downs, the innovation of new instrumentation, like the Optima AUC, improved methodology, and several advances in data analysis technologies [5-9] have reinvigorated interest in AUC [10], resulting in it being one of the leading methods used in biophysical characterization studies. As a result, AUC is used in drug candidate selection, biosimilar characterization, aggregation analysis, formulation studies, and discoveries in gene therapeutics, nanoparticles, proteins, and biologics [3, 11-16].

AUC measures several hydrodynamic properties of homogeneous and heterogeneous samples with high statistical relevance, including molar mass, anisotropy, frictional properties, and partial specific volumes. AUC monitors a sample's sedimentation pattern during centrifugation at user-specified wavelengths. The sample’s sedimentation and diffusion patterns form a moving boundary throughout the experiment (Figure 1.1). After the experiment is completed, the scientist than analyzes the data by fitting and refining a model to the dataset. The analysis removes time- and radially-invariant noise and iteratively refines the models; the refinement steps used in this thesis are described in Table 1.1.

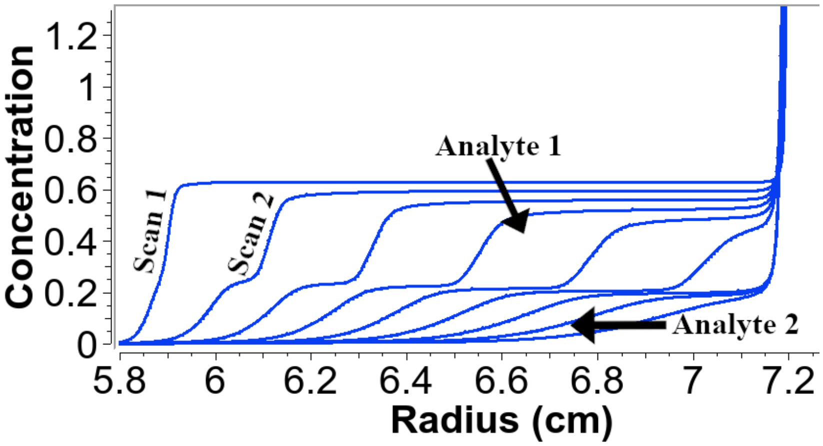


Figure 1.1: Simulated AUC experiment showing the sample boundary progression throughout an experiment. Each line represents the sample boundary at each scan. From inspecting the boundary, it is clear that this sample contains two analytes.

Table 1.1: 2-Dimensional Spectrum Analysis Steps for UltraScan III

Fitting name [a]	**Description**
2DSA	In this first step, a grid is developed to fit the sedimentation and diffusion coefficients. Typically, this is done by parameterizing the diffusion coefficient using the frictional ratio while determining the time-invariant noise. This grid should be large enough to cover all analytes present in the sample, and as the ranges are increased, the grid resolution should also be increased. This step creates the first model for our data set.
2DSA Fit meniscus and/or fit bottom (2DSA-FM / 2DSA-FM/B)	Once a grid that fits the data well has been found, the same grid is used to determine the time- and radially-invariant noise profiles and to fit the meniscus (FM) and/or bottom of the cell (FMB/FB). A fit meniscus should be performed for all samples because the user or software initially visually selects the meniscus position, therefore potentially introducing some error. The fit meniscus generates 11 models, one at the meniscus position selected and 5 to the left and the right of the selected meniscus, within a 0.03 cm range (the user can change this value). After the models are generated, the user should use the Fit Mensicus program to select the model with the lowest RMSD, and all other models can be deleted. The same process can be done for the bottom of the cell, which is highly recommended when working with floating or highly diffusing samples.
2DSA iterative (2DSA-IT)	After the 2DSA-FM/FB is completed, the same grid is used again to further refine the model by performing up to 10 iterative fits, using both time- and radially-invariant noise determination. The produced model at this time should represent the raw data accurately, which is determined based on the quality of the fit [a]. Further analysis steps can be used to refine the model if wanted. They can be found at https://www.ultrascan3.aucsolutions.com/sed-veloc-flowchart.php

[a] Steps should be performed in the order listed. However, after each step, the accuracy of the model should be investigated by checking how well the model overlays the raw data, the RMSD, and the residual pixel map. If the fit is poor, the fitting parameters should be inspected and modified, and the previous step repeated.

The first ultracentrifuge instrument was developed by Theodor Svedberg in the 1920s and was equipped with an optical system to gather sedimentation data [17, 18]. The protocol developed by Svedberg is remarkably similar to the AUCs used today. In honour of his work, the unit Svedberg (S) was named after him. In addition, Svedberg's Ph.D. student, O. Lamm, derived the Lamm equation [19], which describes the space-time evolution of the solutes concentrations in terms of sedimentation (*S*) and diffusion (*D*) under centrifugal force in a sector-shaped cell:

$$\frac{\partial c}{\partial t} = D\left[\left(\frac{\partial^2 c}{\partial r^2}\right) + \frac{1}{r}\left(\frac{\partial c}{\partial r}\right)\right] - S\omega^2\left[r\left(\frac{\partial c}{\partial r}\right) + 2c\right]$$

(Equation 1.1)

where c is the solute concentration, ω is the rotor's angular velocity, and t and r are the time and radius from the center of the rotor, respectively. This equation describes the competing transport processes of sedimentation, which concentrates the sample towards the bottom of the cell under centrifugal forces, and diffusion, which seeks to equalize the solute concentration throughout the solution. For sedimentation velocity (SV) experiments, the sedimentation coefficient (*S*) of a macromolecule is determined using the Svedberg equation:

$$S = \frac{u}{\omega^2 r} = \frac{M(1 - \bar{v}\rho_{sol})}{N_A f}$$

(Equation 1.2)

where u is the velocity of the macromolecule. The sedimentation of the molecule also depends on the solute and solvent parameters, as seen in the second half of the equation, where M is the molar mass of the solute, $\bar{v}$ is the partial specific volume of the solute, which is the inverse of the solute's density and includes its hydration sphere, f is the frictional coefficient of the solute, ρ_{sol} is the density of the solvent, and N_A is Avogadro's number.

Diffusion results in the radial broadening of the sedimentation boundary during the experiment and can be determined by:

$$D = \frac{RT}{N_A f}$$

(Equation 1.3)

where R is the gas constant, and T is the temperature. Further, the anisotropy of the molecule can be determined by dividing the frictional coefficient (f) by the frictional coefficient of its minimal sphere (f_0), obtained by the Stokes-Einstein relationship [20]:

$$f_0 = 6\pi\eta r_o$$

(Equation 1.4)

where η is the viscosity of the buffer, and r_0 is the radius of the minimal sphere of the molecule based on its volume (V).

$$r_0 = \left(\frac{3V}{4\pi}\right)^{1/3}$$

(Equation 1.5)

Understanding a solution's sedimentation and diffusion coefficient profile provides information on its hydrodynamic, thermodynamic, and density properties. Various AUC methods can be implemented depending on the experimental goals and sample [21, 22]. This thesis will focus on three methods utilizing multi-wavelength (MW) optics to characterize gene therapeutics, focusing on their experimental design and analysis.

1.2 Multi-wavelength AUC

Technologies such as the Optima AUC from Beckman Coulter and a system from the Cölfen lab, which is part of the OpenAUC project, have made multi-wavelength AUC (MW-AUC) possible [23]. Each system has its strengths, and a comparison of the two systems can be found in Henrickson A., et al. [13]. Single-wavelength SV experiments are great for determining the hydrodynamic properties of pure samples. However, when measuring a sample containing multiple analytes with different extinction spectra, MW-AUC can elucidate each analyte's concentration, sedimentation coefficient, and binding ratio if the molar extinction coefficients are known. MW-AUC has revolutionized the study of protein-nucleic acid interactions [12, 24, 25], nanoparticles [16, 26, 27], and gene therapy formulations [14, 28]. It measures a sample over a broad wavelength range and utilizes an analyte's unique absorbance spectrum, which can be intrinsic to the molecule or extrinsic from an added fluorescent tag or dye, to differentiate it from other analytes in the solution [13]. When analytes with different spectra are combined, the resulting spectrum is typically an additive combination of the original spectra. However, in some cases, interactions between the analytes can change their original spectrum [29], in which case the quality of the optical deconvolution, described below, would be impacted.

A MW-AUC experiment monitors a sample's sedimentation pattern at multiple wavelengths, usually between 15-30 wavelengths, throughout the experiment. The MW data can be visualized as a 3D or 4D plot to help identify which absorbing analytes are present in each hydrodynamically separated species (Figure 1.2). If each analyte's pure spectral profiles in the solution are known, they can be deconvoluted into the MW data at each radial and time point [9], producing an AUC dataset for each absorbing analyte (Figure 1.2). If the molar extinction coefficients of the analytes are known, then the deconvolution can produce the molar concentrations of each analyte in the hydrodynamically separated solutes, and a binding ratio can be determined [12, 25]. The optically deconvoluted AUC data set for each absorbing analyte can then be analyzed [30] and overlayed to identify what percentage of the analytes are free in solution and what percentage are interacting.

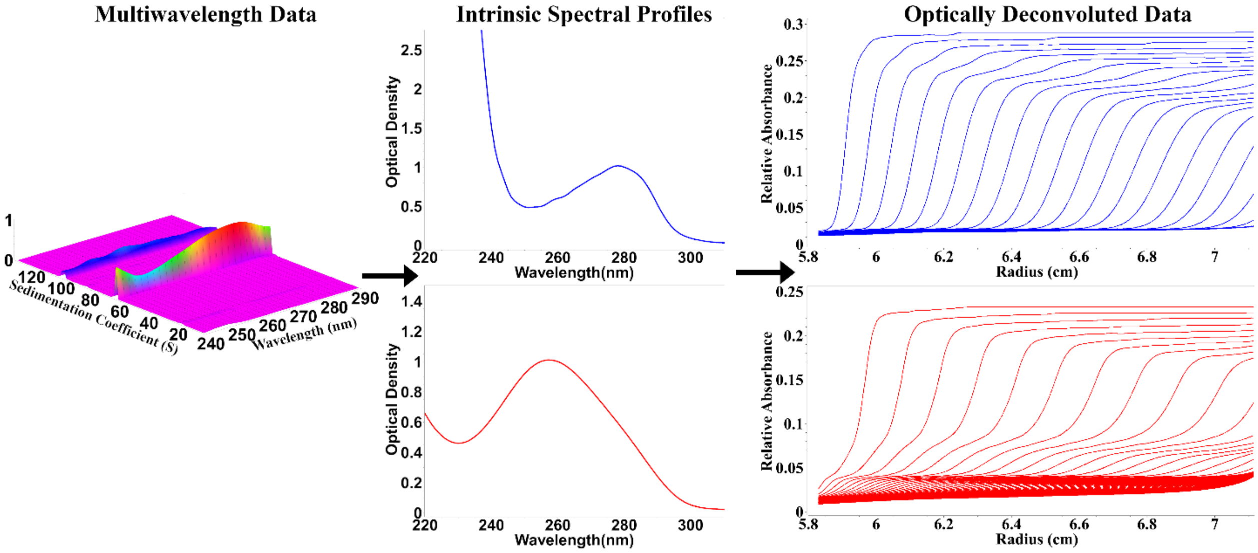


Figure 1.2: Multi-wavelength AUC workflow: The left panel shows a 3D plot of the multiwavelength data, which is optically deconvoluted into two pure spectral profiles of the absorbing analytes, protein (blue) and DNA (red) (center panel), resulting in the optically deconvoluted hydrodynamic data sets on the right.

Typically, 15-30 different wavelengths are collected when measuring in MW mode, and each wavelength should have at least 17 scans to ensure proper noise fitting. However, Beckman Coulter's Optima AUC collects the wavelengths in sequential order, with a minimal scanning time of 16 seconds per scan, depending on the rotor speed. Therefore, at this scanning speed, it is typically only possible to collect enough scans per wavelength when measuring one cell containing up to two samples. However, even with this limitation, the ability of MW to optically deconvolute the analytes provides the user with information that is not readily available from standard single-wavelength AUC experiments, such as binding ratios and accurate percentages of each hydrodynamic species, making MW-AUC a powerful technique [31].

1.3 Density Matching AUC

Density matching AUC has been utilized for decades using isotopic variations of H_2O as well as sucrose, glycerol, and Nycodenz solutions to study membrane proteins in solubilized detergents and to determine the apparent partial specific volumes of proteins [32-35]. The partial specific volume is defined as the change in the volume of a solution as a function of the analyte's concentration and includes its hydration shell. The hydration shell, however, depends on the ionic strength and pressure in which it is measured.

When performing density matching AUC, the sample is measured several times at a constant concentration in buffers that only vary in their hydrogen isotope level, achieved by varying the H_2O:D_2O ratio. The variation in the viscosity and density of the buffers affects the analytes' sedimentation in solution (Figure 1.3 A and B) [36-38]. From the effects measured on the sedimentation coefficient, the analytes' density can be determined by plotting the sedimentation coefficient variations against the buffer density (Figure 1.3 C). Then, for each boundary fraction, the best-fit line can be extrapolated to zero sedimentation, where the analyte's density will equal the buffer's density, and the partial specific volume of the analyte can be determined by taking the inverse of the density [14]. When repeated for every boundary fraction, the results can be plotted as boundary fraction vs. partial specific volume, resulting in a partial specific volume distribution that can be used to identify the partial specific volume for each analyte in the solution (Figure 1.2D). After the partial specific volume is determined, the hydrodynamic radius and molar mass distributions can also be calculated [14]. When using D_2O, one must be aware of the possibility of partial hydrogen-deuterium exchange (HDX), which can artificially lower the partial specific volume but can be avoided when using $H_2{}^{18}O$ [39].

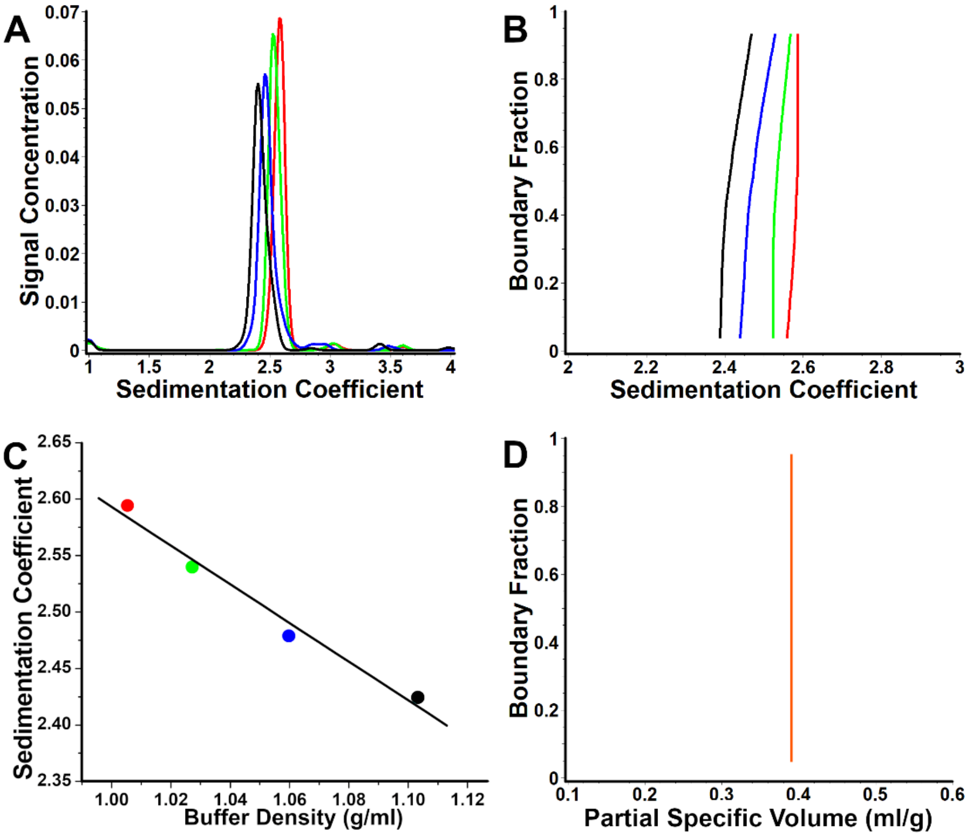


Figure 1.3: Density Matching AUC workflow for a 21 nt siRNA: siRNA was sedimented in buffer containing 0% D_2O (red), 20 % D_2O (green), 50% D_2O (blue), and 90% D_2O (black). The g(s) plots are shown in A, and the integral plots in B. C is produced by plotting the sedimentation coefficient against the density of the buffer, this needs to be repeated for each boundary fraction. The best-fit line between the points, at each boundary fraction, is extrapolated to 0 S and plotted as boundary fraction vs. partial specific volume (the inverse of density), as seen in D.

Density matching AUC is now beginning to enter the gene therapeutics field for LNPs and AAVs [14, 15, 28]. It is a powerful method for LNPs because the lipid's density is close to water's density, resulting in significant sedimentation shifts when the buffer's isotope levels vary. This method also allows the researcher to determine the loaded RNA copy number in their LNPs if the partial specific volume of the empty LNPs is known, helping them to understand the therapeutics payload distribution [15].

1.4 Analytical Buoyant Density Equilibrium AUC

ABDE experiments are performed at equilibrium, differentiating them from the sedimentation velocity experiments explained above. This equilibrium state is achieved using a density gradient forming material, which establishes a density gradient when centrifugal forces are applied. CsCl is an ideal option for absorption studies as it has minimal absorption above 240 nm, and the concentration can be modulated to separate species in a solution based on their densities [40]. When a centrifugal force is applied, the density gradient begins to form, and the solutes either sediment or float to their isopycnic point within the gradient, forming a peak at this position. The sample concentration determines the peak’s intensity, and the analyte's diffusion determines the peak's broadness. Meselson and Stahl pioneered density gradient studies in the AUC in 1958 while elucidating the semi-conservative replication of DNA [41]. Today, density gradient studies are undergoing a renaissance in the field of gene therapies [40, 42-44]. Viral gene therapies benefit significantly from this method due to the density differences between full and empty capsids [43]; even partially filled capsids can be identified in the gradients [40]. This level of separation can be challenging to achieve with many other techniques as the capsids do not differ in their hydrodynamic radii.

An advantage of measuring at an equilibrium state, as in ABDE experiments, is that up to 100 wavelengths can be collected in MW-AUC. Unlike SV experiments, where the number of wavelengths measured is constrained to allow enough scans to be collected before the sample has fully sedimented, ABDE experiments do not have this limitation and can still be optically deconvoluted into their spectral contributors [40]. Another advantage of measuring at equilibrium is that more cells can be measured in a single experiment in MW mode, increasing the throughput. ABDE methods also require significantly less sample than velocity experiments, as the sample concentrates at its isopycnic point within the cell, and the sample requirements can be further reduced by changing the centerpiece's pathlength from 1.2 cm to 0.3 cm [40].

Although these data can be analyzed in spreadsheet programs such as OriginPro2022 or Microsoft Excel, UltraScan has developed two new software modules dedicated to this analysis, allowing for easy optical deconvolution, visualization, and analysis of the MW experiment [7, 45]. The first module, Pseudo-absorbance, allows the user to remove systemic noises resulting from the AUC, such as time and radially-invariant noises, without subtracting a reference channel [7], improving the signal-to-noise ratio and increasing the number of channels available for sample measurements. However, if the density gradient forming material used absorbs at the wavelengths measured, a single buffer channel can be measured, and the baseline can be subtracted from all channels in that experiment using this program. The second program, ABDE analysis, can determine each peak's radial point, density, and sample percentage [45]. The program can also overlay different ABDE experiments and normalize them, if needed, to help present the data. It should be noted that the analyte's apparent partial specific volume determined in a density gradient forming material has been found to differ from those calculated in aqueous solutions and is generally determined to be smaller due to interactions between the analyte and CsCl [44].

1.5 Gene Therapeutics

Gene therapy is the transfer of genetic material to a patient to treat or cure a disease. It has been around for decades, with clinical trials starting in the 1990s, and is now a highly studied area with over 2500 clinical studies and, as of May 2023, 29 approved cell and gene therapy treatments [46]. The genetic material can consist of DNA, mRNA, siRNA, or antisense oligonucleotides. This genetic material is loaded into vectors, which helps direct it to specific organs, protect it from degradation, and reduce adverse health effects [47, 48]. These vectors can be non-viral, including liposomes, polymers, and nanoparticles [49], or viral-based, including adenovirus, lentivirus, retrovirus, and adeno-associated virus (AAV) [50]. The most actively studied viral and non-viral vectors are AAVs and lipid nanoparticles (LNPs), respectively [48, 51].

Gene therapy development processes have faced several hurdles, some of which are related to manufacturing capacity, limited characterization methods, and immune responses [52]. The characterization methods need to be able to identify several aspects of the therapeutic to ensure it passes quality control requirements. These requirements include product testing to ensure no microbial contaminants are present, quantifying and controlling process- and product-related impurities such as empty vectors, potency determination, and stability testing [52]. Characterization methods that can accurately assess these requirements need to be used to ensure patient safety and help inform the development and purification pathways.

1.5.1 Lipid Nanoparticles

LNPs are currently regarded as the most promising non-viral vector delivery system [48]. They can package a variety of nucleic acids, including siRNA, mRNA, and plasmid DNA, as well as other small molecules [48]. LNP-based therapies allow therapeutic delivery *in vivo* and aim to elicit an effective adaptive immune response that generates an immunogenic memory, so the body can produce antibodies faster when an infection occurs, as seen with the COVID-19 vaccines [53, 54]. LNPs are spherical vesicles made of ionizable lipids, cholesterol, and polyethylene glycol (PEG), used to extend the circulation time in the blood. The formulations are manufactured through rapid mixing techniques [55]. The lipid type, size, and charge impact the

LNP behaviour *in vivo* [56]; therefore, the lipid components and mixing techniques must be optimized to obtain reproducible and scalable LNPs.

The LNP formulations need to pass several quality tests, including evaluation of physical characteristics such as hydrodynamic radius, RNA encapsulation and loading number, and charge [57]. DLS is a widely used technique for size determination. However, it is inappropriate for highly polydisperse samples, such as LNP preparations. Because the larger particles result in Mie scattering, which scales to the sixth power of the molecule’s radius, potentially resulting in a masking of the smaller particles and an incorrect size distribution [58]. Due to the intrinsic complexity of LNPs, characterization is challenging. However, improved methodology in density matching AUC could assist in the characterization of LNPs. Density matching AUC can provide information on the LNP particle size distributions, partial specific volume, and nucleic acid loading number in LNPs [14, 15].

1.5.2 Adeno-Associated Virus

AAVs are small, naturally occurring, non-enveloped viruses, about 25 nm in diameter, that package single-stranded DNA and are part of the *Parvoviridae* family [59, 60]. They are not known to cause disease and cannot propagate without a helper virus, making them a strong gene therapy candidate [51]. They are used for *in vivo* gene delivery and can transduce several tissues, depending on their serotype, including the liver, muscle, and nervous tissues [61]. They are generally used when long-term transgene expression and disease correction is wanted [61, 62].

The AAV capsid is made up of the three proteins VP1, VP2, and VP3, which are typically found in a 5:5:50 ratio [63]. Recombinant AAVs are produced by replacing the AAV genome with the transgene of interest between the two inverted terminal repeats (ITRs). The ITRs act as the primary packaging signal, allowing the DNA to enter the capsid [51, 64].

Typically, recombinant AAVs are produced by transfecting HEK 293 cells with three plasmids: one containing the transgene flanked by the ITRs, one encoding the AAV capsid proteins, and one encoding helper proteins required to assemble the capsid [65]. Initially, empty AAV capsids are produced, then, through interactions with the ITR regions, a single strand of the transgene is packaged into the capsid. Consequently, this method results in the production of loaded AAVs as well as empty AAVs, a product-related impurity. Other product-related impurities, such as partially filled and over-filled AAVs, can also be produced.

The empty AAVs may not be directly harmful to the patient but would likely increase the immune response, while the partially filled and overfilled AAVs may contain portions of the transgene or host DNA, which could negatively impact the patient if expressed. Detecting these impurities is challenging when using sizing techniques such as SEC, AF4, and DLS, as the product-related impurities have similar hydrodynamic radii and surface properties compared to the full AAVs. Currently, AUC is the only established technique that can characterize all these potential impurities in a sample [66, 67].

AUC is a powerful technique for which many methods are available, depending on the sample and research goals. This thesis aims to develop upon MW-AUC and demonstrate feasible experimental designs that researchers can implement in their studies by showing a few examples and considerations of MW SV experiments. It then applies MW to density matching and ABDE experiments to characterize gene therapy formulations, including LNPs and AAVs. Implementing AUC as a characterization technique for interacting systems and gene therapeutics can help improve vaccine performance and reduce their development times, a critical aspect when responding to emerging pathogen outbreaks.

1.6 References

1. Gandhi AV, Pothecary MR, Bain DL, Carpenter JF. Some lessons learned from a comparison between sedimentation velocity analytical ultracentrifugation and size exclusion chromatography to characterize and quantify protein aggregates. J Pharm Sci. 2017;106(8):2178-2186.

2. Kaur H. Stability testing in monoclonal antibodies. Crit Rev Biotechnol. 2021;41(5):692-714.

3. Berkowitz SA. Role of analytical ultracentrifugation in assessing the aggregation of protein biopharmaceuticals. The AAPS journal. 2006;8:E590-E605.

4. Yu LT, Hancu MC, Kreutzberger MAB, Henrickson A, Demeler B, Egelman EH, et al. Hollow Octadecameric Self-Assembly of Collagen-like Peptides. J Am Chem Soc. 2023;145(9):5285-5296. Epub 2023/02/23. doi: 10.1021/jacs.2c12931. PubMed PMID: 36812303; PubMed Central PMCID: PMCPMC10131286.

5. Walter J, Sherwood PJ, Lin W, Segets D, Stafford WF, Peukert W. Simultaneous analysis of hydrodynamic and optical properties using analytical ultracentrifugation equipped with multiwavelength detection. Anal Chem. 2015;87(6):3396-3403. Epub 2015/02/14. doi: 10.1021/ac504649c. PubMed PMID: 25679871.

6. Philo JS. SEDNTERP: a calculation and database utility to aid interpretation of analytical ultracentrifugation and light scattering data. Eur Biophys J. 2023. Epub 2023/02/16. doi: 10.1007/s00249-023-01629-0. PubMed PMID: 36792822.

7. Mortezazadeh S, Demeler B. Systematic noise removal from analytical ultracentrifugation data with UltraScan. Eur Biophys J. 2023. Epub 2023/02/15. doi: 10.1007/s00249-023-01631-6. PubMed PMID: 36786920.

8. Savelyev A, Gorbet GE, Henrickson A, Demeler B. Moving analytical ultracentrifugation software to a good manufacturing practices (GMP) environment. PLoS Comput Biol. 2020;16(6):e1007942. doi: 10.1371/journal.pcbi.1007942.

9. Mortezazadeh S, Demeler B. A spectral decomposition quality assessment tool for multi-wavelength AUC experiments with UltraScan. Eur Biophys J. 2023. Epub 2023/03/18. doi: 10.1007/s00249-023-01640-5. PubMed PMID: 36930298.

10. Dobson RCJ, Patel TR. Analytical ultracentrifugation: still the gold standard that offers multiple solutions. Eur Biophys J. 2020;49(8):673-676. Epub 2020/11/20. doi: 10.1007/s00249-020-01483-4. PubMed PMID: 33211149.

11. Liu J, Yadav S, Andya J, Demeule B, Shire SJ. Analytical Ultracentrifugation and Its Role in Development and Research of Therapeutical Proteins. Methods Enzymol. 2015;562:441-476. Epub 2015/09/29. doi: 10.1016/bs.mie.2015.04.008. PubMed PMID: 26412663.

12. Ahmed I, Hahn J, Henrickson A, Khaja FT, Demeler B, Dubnau D, et al. Structure-function studies reveal ComEA contains an oligomerization domain essential for transformation in gram-positive bacteria. Nat Commun. 2022;13(1):7724. Epub 2022/12/14. doi: 10.1038/s41467-022-35129-0. PubMed PMID: 36513643; PubMed Central PMCID: PMCPMC9747964.

13. Henrickson A, Gorbet GE, Savelyev A, Kim M, Hargreaves J, Schultz SK, et al. Multi-wavelength analytical ultracentrifugation of biopolymer mixtures and interactions. Anal Biochem. 2022:114728.

14. Henrickson A, Kulkarni JA, Zaifman J, Gorbet GE, Cullis PR, Demeler B. Density matching multi-wavelength analytical ultracentrifugation to measure drug loading of lipid nanoparticle formulations. ACS nano. 2021;15(3):5068-5076.

15. Bepperling A, Richter G. Determination of mRNA copy number in degradable lipid nanoparticles via density contrast analytical ultracentrifugation. Eur Biophys J. 2023. Epub 2023/06/08. doi: 10.1007/s00249-023-01663-y. PubMed PMID: 37289289.

16. Walter J, Löhr K, Karabudak E, Reis W, Mikhael J, Peukert W, et al. Multidimensional analysis of nanoparticles with highly disperse properties using multiwavelength analytical ultracentrifugation. ACS Nano. 2014;8(9):8871-8886. Epub 2014/08/19. doi: 10.1021/nn503205k. PubMed PMID: 25130765.

17. Svedberg T, Pedersen KO. The ultracentrifuge. The Ultracentrifuge. 1940.

18. Kyle RA, Shampo MA. Theodor Svedberg and the ultracentrifuge. Mayo Clin Proc. 1997;72(9):830. Epub 1997/09/19. doi: 10.4065/72.9.830. PubMed PMID: 9294529.

19. Lamm O. Die differentialgleichung der ultrazentrifugierung: Almqvist & Wiksell; 1929.

20. Einstein A. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. Annalen der physik. 1905;4.

21. Schuck P, Zhao H, Brautigam CA, Ghirlando R. Basic principles of analytical ultracentrifugation: CRC Press; 2016.

22. Aziz Z, Daugherty M, de la Torre J, Demeler B, Douady C, Durchschlag H, et al. Analytical ultracentrifugation: techniques and methods: Royal Society of Chemistry; 2007.

23. Pearson J, Walter J, Peukert W, Cölfen H. Advanced Multiwavelength Detection in Analytical Ultracentrifugation. Anal Chem. 2018;90(2):1280-1291. Epub 2017/12/08. doi: 10.1021/acs.analchem.7b04056. PubMed PMID: 29214799.

24. Horne CR, Henrickson A, Demeler B, Dobson RC. Multi-wavelength analytical ultracentrifugation as a tool to characterise protein–DNA interactions in solution. Eur Biophys J. 2020;49(8):819-827.

25. Horne CR, Venugopal H, Panjikar S, Wood DM, Henrickson A, Brookes E, et al. Mechanism of NanR gene repression and allosteric induction of bacterial sialic acid metabolism. Nat Commun. 2021;12(1):1988. Epub 2021/04/02. doi: 10.1038/s41467-021-22253-6. PubMed PMID: 33790291; PubMed Central PMCID: PMCPMC8012715.

26. Wawra SE, Pflug L, Thajudeen T, Kryschi C, Stingl M, Peukert W. Determination of the two-dimensional distributions of gold nanorods by multiwavelength analytical ultracentrifugation. Nat Commun. 2018;9(1):4898. Epub 2018/11/23. doi: 10.1038/s41467-018-07366-9. PubMed PMID: 30464237; PubMed Central PMCID: PMCPMC6249260.

27. Schneider CM, Cölfen H. High-Resolution Analysis of Small Silver Clusters by Analytical Ultracentrifugation. J Phys Chem Lett. 2019;10(21):6558-6564. Epub 2019/10/09. doi: 10.1021/acs.jpclett.9b02755. PubMed PMID: 31592670.

28. Maruno T, Usami K, Ishii K, Torisu T, Uchiyama S. Comprehensive size distribution and composition analysis of adeno-associated virus vector by multiwavelength sedimentation velocity analytical ultracentrifugation. J Pharm Sci. 2021;110(10):3375-3384.

29. Gosetti F, Mazzucco E, Zampieri D, Gennaro MC. Signal suppression/enhancement in high-performance liquid chromatography tandem mass spectrometry. J Chromatogr A. 2010;1217(25):3929-3937. Epub 2009/12/17. doi: 10.1016/j.chroma.2009.11.060. PubMed PMID: 20004403.

30. Brookes E, Cao W, Demeler B. A two-dimensional spectrum analysis for sedimentation velocity experiments of mixtures with heterogeneity in molecular weight and shape. Eur Biophys J. 2010;39(3):404-414.

31. Johnson CN, Gorbet GE, Ramsower H, Urquidi J, Brancaleon L, Demeler B. Multi-wavelength analytical ultracentrifugation of human serum albumin complexed with porphyrin. Eur Biophys J. 2018;47(7):789-797.

32. Mayer G, Ludwig B, Müller HW, van den Broek JA, Friesen RHE, Schubert D, editors. Studying membrane proteins in detergent solution by analytical ultracentrifugation: different methods for density matching. Analytical Ultracentrifugation V; 1999 1999//; Berlin, Heidelberg: Springer Berlin Heidelberg.

33. Schubert D, Schuck P, editors. Analytical ultracentrifugation as a tool for studying membrane proteins. Progress in Analytical Ultracentrifugation; 1991 1991//; Darmstadt: Steinkopff.

34. Reynolds JA, Tanford C. Determination of molecular weight of the protein moiety in protein-detergent complexes without direct knowledge of detergent binding. Proc Natl Acad Sci U S A. 1976;73(12):4467-4470. Epub 1976/12/01. doi: 10.1073/pnas.73.12.4467. PubMed PMID: 188041; PubMed Central PMCID: PMCPMC431502.

35. Lustig A, Engel A, Tsiotis G, Landau EM, Baschong W. Molecular weight determination of membrane proteins by sedimentation equilibrium at the sucrose or nycodenz-adjusted density of the hydrated detergent micelle. Biochim Biophys Acta. 2000;1464(2):199-206. Epub 2000/03/23. doi: 10.1016/s0005-2736(99)00254-0. PubMed PMID: 10727607.

36. Brown PH, Balbo A, Zhao H, Ebel C, Schuck P. Density contrast sedimentation velocity for the determination of protein partial-specific volumes. PLoS ONE. 2011;6(10):e26221. Epub 2011/10/27. doi: 10.1371/journal.pone.0026221. PubMed PMID: 22028836; PubMed Central PMCID: PMCPMC3197611.

37. Cheng P, Schachman H. Studies on the validity of the Einstein viscosity law and Stokes' law of sedimentation. Journal of polymer science. 1955;16(81):19-30.

38. Martin W, Cook W, Winkler C. The determination of partial specific volumes by differential sedimentation. Canadian Journal of Chemistry. 1956;34(6):809-814.

39. Englander SW, Sosnick TR, Englander JJ, Mayne L. Mechanisms and uses of hydrogen exchange. Curr Opin Struct Biol. 1996;6(1):18-23. Epub 1996/02/01. doi: 10.1016/s0959-440x(96)80090-x. PubMed PMID: 8696968; PubMed Central PMCID: PMCPMC3412065.

40. Henrickson A, Ding X, Seal AG, Qu Z, Tomlinson L, Forsey J, et al. Characterization and quantification of AAV capsid loading states by multi-wavelength AUC with UltraScan. Nanomedicine. 2023;Submitted.

41. Meselson M, Stahl FW, Vinograd J. EQUILIBRIUM SEDIMENTATION OF MACROMOLECULES IN DENSITY GRADIENTS. Proc Natl Acad Sci U S A. 1957;43(7):581-588. Epub 1957/07/15. doi: 10.1073/pnas.43.7.581. PubMed PMID: 16590059; PubMed Central PMCID: PMCPMC528502.

42. Yang X, Agarwala S, Ravindran S, Vellekamp G. Determination of particle heterogeneity and stability of recombinant adenovirus by analytical ultracentrifugation in CsCl gradients. J Pharm Sci. 2008;97(2):746-763. Epub 2007/06/27. doi: 10.1002/jps.21008. PubMed PMID: 17593554.

43. Berkowitz SA, Philo JS. Monitoring the homogeneity of adenovirus preparations (a gene therapy delivery system) using analytical ultracentrifugation. Anal Biochem. 2007;362(1):16-37.

44. Sternisha SM, Wilson AD, Bouda E, Bhattacharya A, VerHeul R. Optimizing high-throughput viral vector characterization with density gradient equilibrium analytical ultracentrifugation. Eur Biophys J. 2023. Epub 2023/05/03. doi: 10.1007/s00249-023-01654-z. PubMed PMID: 37130969.

45. Savelyev A, Brookes EH, Henrickson A, Demeler B. A new UltraScan module for the characterization and quantification of analytical buoyant density equilibrium experiments to determine AAV capsid loading. Eur Biophys J. 2023. Epub 2023/04/05. doi: 10.1007/s00249-023-01641-4. PubMed PMID: 37014454.

46. FDA. Approved Cellular and Gene Therapy Products [updated May 2023]. Available from: https://www.fda.gov/vaccines-blood-biologics/cellular-gene-therapy-products/approved-cellular-and-gene-therapy-products.

47. Korneyenkov MA, Zamyatnin AA, Jr. Next Step in Gene Delivery: Modern Approaches and Further Perspectives of AAV Tropism Modification. Pharmaceutics. 2021;13(5). Epub 2021/06/03. doi: 10.3390/pharmaceutics13050750. PubMed PMID: 34069541; PubMed Central PMCID: PMCPMC8160765.

48. Yan Y, Liu XY, Lu A, Wang XY, Jiang LX, Wang JC. Non-viral vectors for RNA delivery. J Control Release. 2022;342:241-279. Epub 2022/01/13. doi: 10.1016/j.jconrel.2022.01.008. PubMed PMID: 35016918; PubMed Central PMCID: PMCPMC8743282.

49. Zu H, Gao D. Non-viral Vectors in Gene Therapy: Recent Development, Challenges, and Prospects. AAPS J. 2021;23(4):78. Epub 2021/06/03. doi: 10.1208/s12248-021-00608-7. PubMed PMID: 34076797; PubMed Central PMCID: PMCPMC8171234.

50. Milone MC, O'Doherty U. Clinical use of lentiviral vectors. Leukemia. 2018;32(7):1529-1541. Epub 2018/04/15. doi: 10.1038/s41375-018-0106-0. PubMed PMID: 29654266; PubMed Central PMCID: PMCPMC6035154 utilizing lentiviral vectors for which he has received royalty payments. UOD declares that she has no conflict of interest.

51. Naso MF, Tomkowicz B, Perry WL, Strohl WR. Adeno-associated virus (AAV) as a vector for gene therapy. BioDrugs. 2017;31(4):317-334.

52. Wright JF. Quality Control Testing, Characterization and Critical Quality Attributes of Adeno-Associated Virus Vectors Used for Human Gene Therapy. Biotechnol J. 2021;16(1):e2000022. Epub 2020/11/05. doi: 10.1002/biot.202000022. PubMed PMID: 33146911.

53. Haas EJ, Angulo FJ, McLaughlin JM, Anis E, Singer SR, Khan F, et al. Impact and effectiveness of mRNA BNT162b2 vaccine against SARS-CoV-2 infections and COVID-19 cases, hospitalisations, and deaths following a nationwide vaccination campaign in Israel: an observational study using national surveillance data. Lancet. 2021;397(10287):1819-1829. Epub 2021/05/09. doi: 10.1016/s0140-6736(21)00947-8. PubMed PMID: 33964222; PubMed Central PMCID: PMCPMC8099315 stock options in Pfizer. All other authors declare no competing interests.

54. Martínez-Baz I, Trobajo-Sanmartín C, Miqueleiz A, Guevara M, Fernández-Huerta M, Burgui C, et al. Product-specific COVID-19 vaccine effectiveness against secondary infection in close contacts, Navarre, Spain, April to August 2021. Euro Surveill. 2021;26(39). Epub 2021/10/02. doi: 10.2807/1560-7917.Es.2021.26.39.2100894. PubMed PMID: 34596016; PubMed Central PMCID: PMCPMC8485582.

55. Belliveau NM, Huft J, Lin PJ, Chen S, Leung AK, Leaver TJ, et al. Microfluidic Synthesis of Highly Potent Limit-size Lipid Nanoparticles for In Vivo Delivery of siRNA. Mol Ther Nucleic Acids. 2012;1(8):e37. Epub 2013/01/25. doi: 10.1038/mtna.2012.28. PubMed PMID: 23344179; PubMed Central PMCID: PMCPMC3442367.

56. Let’s talk about lipid nanoparticles. Nature Reviews Materials. 2021;6(2):99-99. doi: 10.1038/s41578-021-00281-4.

57. Hou X, Zaks T, Langer R, Dong Y. Lipid nanoparticles for mRNA delivery. Nature Reviews Materials. 2021;6(12):1078-1094. doi: 10.1038/s41578-021-00358-0.

58. Tomaszewska E, Soliwoda K, Kadziola K, Tkacz-Szczesna B, Celichowski G, Cichomski M, et al. Detection Limits of DLS and UV-Vis Spectroscopy in Characterization of Polydisperse Nanoparticles Colloids. J Nanomaterials. 2013;2013. doi: 10.1155/2013/313081. PubMed PMID: 10.1155/2013/313081.

59. Cotmore SF, Agbandje-McKenna M, Canuti M, Chiorini JA, Eis-Hubinger AM, Hughes J, et al. ICTV Virus Taxonomy Profile: Parvoviridae. J Gen Virol. 2019;100(3):367-368. Epub 2019/01/24. doi: 10.1099/jgv.0.001212. PubMed PMID: 30672729; PubMed Central PMCID: PMCPMC6537627.

60. Wang XS, Ponnazhagan S, Srivastava A. Rescue and replication signals of the adeno-associated virus 2 genome. J Mol Biol. 1995;250(5):573-580. Epub 1995/07/28. doi: 10.1006/jmbi.1995.0398. PubMed PMID: 7623375.

61. Mingozzi F, High KA. Therapeutic in vivo gene transfer for genetic disease using AAV: progress and challenges. Nat Rev Genet. 2011;12(5):341-355. Epub 2011/04/19. doi: 10.1038/nrg2988. PubMed PMID: 21499295.

62. Murillo O, Luqui DM, Gazquez C, Martinez-Espartosa D, Navarro-Blasco I, Monreal JI, et al. Long-term metabolic correction of Wilson's disease in a murine model by gene therapy. J Hepatol. 2016;64(2):419-426. Epub 2015/09/27. doi: 10.1016/j.jhep.2015.09.014. PubMed PMID: 26409215.

63. Wörner TP, Bennett A, Habka S, Snijder J, Friese O, Powers T, et al. Adeno-associated virus capsid assembly is divergent and stochastic. Nat Commun. 2021;12(1):1642. Epub 2021/03/14. doi: 10.1038/s41467-021-21935-5. PubMed PMID: 33712599; PubMed Central PMCID: PMCPMC7955066 in employing AAV vectors for gene delivery purposes. M.A.-M. is co-founder of StrideBio, Inc., with an interest in developing AAV technology for gene delivery purposes. M.A.-M. is a member of the ATGC and Voyager SAB, and consultants for StrideBio, Intima Bioscience, being biopharma companies with interest in developing AAV for gene delivery purposes. The remaining authors declare no competing interests.

64. Samulski RJ, Muzyczka N. AAV-Mediated Gene Therapy for Research and Therapeutic Purposes. Annu Rev Virol. 2014;1(1):427-451. Epub 2014/11/03. doi: 10.1146/annurev-virology-031413-085355. PubMed PMID: 26958729.

65. Challis RC, Kumar SR, Chan KY, Challis C, Beadle K, Jang MJ, et al. Systemic AAV vectors for widespread and targeted gene delivery in rodents. Nat Protoc. 2019;14(2):379-414.

66. Burnham B, Nass S, Kong E, Mattingly M, Woodcock D, Song A, et al. Analytical ultracentrifugation as an approach to characterize recombinant adeno-associated viral vectors. Human gene therapy methods. 2015;26(6):228-242.

67. Fu X, Chen WC, Argento C, Clarner P, Bhatt V, Dickerson R, et al. Analytical Strategies for Quantification of Adeno-Associated Virus Empty Capsids to Support Process Development. Hum Gene Ther Methods. 2019;30(4):144-152. Epub 2019/08/02. doi: 10.1089/hgtb.2019.088. PubMed PMID: 31368356.

Chapter 2. Multi-wavelength analytical ultracentrifugation of biopolymer mixtures and interactions

Amy Henrickson[1], Gary E. Gorbet[2], Alexey Savelyev[3], Minji Kim[4], Jason Hargreaves[5], Sarah K. Schultz[1], Ute Kothe[1,6], Borries Demeler[1,2,3]

[1] University of Lethbridge, Dept. of Chemistry and Biochemistry, Lethbridge, Alberta, Canada
[2] AUC Solutions, LLC, Houston, TX, USA
[3] University of Montana, Dept. of Chemistry, Missoula, MT, USA
[4] Carnegie Mellon University, Dept. of Computer Science, Pittsburgh, PA, USA
[5] Botaneco INC, Calgary, Alberta, Canada
[6] University of Manitoba, Dept. of Chemistry, Winnipeg, Manitoba, Canada

2.1 Contributions of Authors

This chapter consists of an adapted manuscript published in *Analytical Biochemistry,* Volume 652, article 114728, on May 26, 2022 (https://doi.org/10.1016/j.ab.2022.114728), prepared in collaboration with Gary E. Gorbet, Dr. Alexey Savelyev, Minji Kim, Dr. Jason Hargreaves, Sarah K. Schultz, Dr. Ute Kothe, and Dr. Borries Demeler. I performed all AUC experiments and analyses and assisted with writing the manuscript with Dr. Borries Demeler, who also conceived the experiments. Gary E. Gorbet, Dr. Alexey Savelyev, and Minji Kim contributed to the multi-wavelength analysis software development, Dr. Jason Hargreaves provided the oil seed protein samples, and Sarah K. Schultz and Dr. Ute Kothe prepared and contributed the fluorescent protein samples and edited the manuscript.

2.2 Abstract

Multi-wavelength analytical ultracentrifugation (MW-AUC) is a recent development made possible by new analytical ultracentrifuge optical systems. MW-AUC extends the basic hydrodynamic information content of AUC thus providing access to a wide range of new applications for biopolymer characterization, and is poised to become an essential analytical tool to study macromolecular interactions. It adds an orthogonal spectral dimension to the traditional hydrodynamic characterization by exploiting unique chromophores in analyte mixtures that may or may not interact. Here, we illustrate the utility of MW-AUC for experimental investigations where the benefit of the added spectral dimension provides critical information that is not accessible, and impossible to resolve with traditional AUC methods. We demonstrate the improvements in resolution and information content obtained by this technique compared to traditional single- or dual-wavelength approaches and discuss experimental design considerations and limitations of the method. We further address the advantages and disadvantages of the two MW optical systems available today, and the differences in data analysis strategies between the two systems.

2.3 Keywords

Multi-wavelength analytical ultracentrifugation, Macromolecular hetero-interactions, UltraScan AUC software, Composition analysis, Spectral decomposition

2.4 Introduction

Interactions between biopolymers are essential for most biological processes, such as metabolic and developmental control and cell regulation [1], and therefore are of great research interest. Common approaches to study macromolecular interactions include affinity chromatography, gel shift assays, co-immunoprecipitation, x-ray crystallography, mass spectrometry, thermophoresis, isothermal titration calorimetry, and surface plasmon resonance. While these methods have achieved many breakthroughs, they also have their limitations. In our work, we focus on and discuss applications of a recently developed alternative called multi-wavelength analytical ultracentrifugation (MW-AUC). This method can be used to study biopolymer interactions in a physiological environment, where ionic strength, pH, and redox potential can be easily controlled. Drugs and other small molecules can be titrated to study their interactions in a reversible environment, and multiple distinct molecules can participate in hetero-interactions. Taking advantage of unique chromophores and their extinction profiles, MW-AUC can extract additional information from the orthogonal spectral dimension. Previously, a variation of this theme proposed a multi-signal analysis by Brautigam et al. [2], which included Rayleigh interference measurements in the AUC to gain orthogonal information.

In 2008, the Cölfen lab introduced the first fiber-based UV/visible multi-wavelength detector for the analytical ultracentrifuge [3], adding an optical characterization dimension to the traditional hydrodynamic separation. This accomplishment added an important method to the toolkit of analytical ultracentrifugation (AUC), further enhancing the potential for discovery through the already capable and time-honored AUC method. This optical system was further improved in 2015 [4], and our laboratory contributed the data analysis framework implemented in UltraScan [5] for data generated by this detector [6]. In 2018, this method was further enhanced by the addition of mirror optics [7] (here referred to as “Cölfen optics”). This design has been successfully employed in multi-wavelength experiments of biopolymers with chromophores in the visible range [8], protein-DNA mixtures [6], and protein-RNA interactions [9]. The Cölfen optics design has been made available under an open-source license [10]; it is intended to be retrofit into a preparative ultracentrifuge sold by Beckman-Coulter. In addition to the multi-wavelength analysis program implemented in UltraScan for the Cölfen optics, analysis of the same data was also integrated into Sedanal [11], developed by Walter Stafford. This program employs the time-derivative method, which fits differences of scans simulated by finite element solutions of the Lamm equation to experimental scan differences. This approach has the benefit of largely eliminating contributions from time-invariant noise, at the expense of amplifying stochastic noise by a factor of $\sqrt{2}$ from the pairwise subtraction of experimental data scans. In contrast to Sedanal, UltraScan avoids the amplification of stochastic noise by directly fitting time- and radially-invariant noise, and by fitting experimental data directly to linear combinations of finite element solutions [12]. In 2016, Beckman-Coulter released a new generation of analytical ultracentrifuges, the Optima AUC™ series. It was equipped with Rayleigh interference optics and multi-wavelength capable UV/visible absorption optics (here referred to as “Beckman optics”), and is currently the only commercially available analytical ultracentrifuge. MW-AUC experiments with biopolymers performed with the Beckman optics are starting to emerge and include studies on heme proteins [13-15], triphenylmethane dyes binding to peptide trimers derived from amyloid-β peptides [16], and protein-DNA interactions [17]. Here, we describe the pros and cons of both optical systems, provide guidance for experimental designs of MW-AUC experiments, consider the limitations of the MW-AUC approach, and illustrate the significant advantages of MW-AUC over traditional single wavelength AUC experiments. We further describe the distinctions in the data analysis algorithms required to correctly evaluate data from the two optical systems, and document the use and implementation of UltraScan analysis modules required to process MW-AUC experiments.

2.4.1 Principles of MW-AUC

Analytical ultracentrifugation is a technique used to measure the partial concentrations, sedimentation coefficients, and the diffusion coefficients of analytes present in colloidal molecular mixtures. From this information, details about the analyte’s size and degree of globularity can be obtained [18]. Detection of the molecules is traditionally performed by scanning the sedimenting sample using single-wavelength absorbance spectroscopy as a function of radius and time. In a MW-AUC experiment, the sedimenting sample is scanned at multiple wavelengths. If the solution contains different analytes, each characterized by a different absorbance spectrum, MW-AUC detection provides a second, orthogonal characterization method by resolving analytes not just by differences in their hydrodynamic properties, but also by the differences in their absorbance spectra. If the intrinsic molar extinction profiles for each pure analyte are known, and they are sufficiently dissimilar, the spectrum of the mixture can be decomposed into the partial absorbance contributions from each analyte, and the molar quantity of each constituent can be determined [9]. Molecules that form complexes will sediment faster than their unbound forms due to the increase in mass of the complex. The stoichiometry and molar ratio of each analyte in the complex can be deduced by integrating the decomposed spectra. This second dimension adds important information to the hydrodynamic properties, extending the value and impact of traditional AUC.

2.4.2 Differences between the two MW-AUC optical systems

While both UV/visible optical systems mentioned above share mirror-based optics and support the acquisition of experimental data at multiple wavelengths, important differences in the two systems affect how data are collected, stored, and analyzed. These differences also determine the types of experiments that can be performed with the instruments. Both optics use a stepping motor to scan the radial domain rapidly. The Cölfen optics employ a data collection system where white light passes through the sample, and then is passed through a slit assembly, coupled to an optical fiber, and then to a diffraction grating. The diffracted light is then imaged on a linear CCD spectrophotometer, producing a wavelength intensity scan with approximately 0.5 nm resolution for each radial position imaged with the device. In the Beckman optics, white light passes over a diffraction grating before passing monochromatic light with 1 nm resolution through the sample. The resulting monochromatic intensity is imaged for each wavelength sequentially on a photomultiplier tube at each radial position in the AUC cell, producing multiple single-wavelength velocity experimental data sets.

2.4.3 Advantages and limitations of each MW-AUC optical system

These fundamentally different optical systems have pros and cons to be considered in the design, application, and analysis of experiments. An important difference between these optical designs is the order in which data are collected. With the Cölfen optics, experimental data from different wavelengths are collected simultaneously, which offers a distinct scanning speed advantage. The Beckman optics employ a photomultiplier tube which scans monochromatic light at a single wavelength, requiring each wavelength to be acquired sequentially. The use of a photomultiplier tube offers distinct dynamic range advantages, especially in the lower UV range, where fiber-based CCD systems suffer from reduced light intensity and therefore lack sensitivity.

This presents a problem for the case of biopolymers (nucleic acids, proteins, lipids, carbohydrates), where detection often relies on the measurement of chromophores that absorb between 210 nm–240 nm (see Fig. SI 2.1). This lack of sensitivity is further exacerbated when buffer components that absorb below 260 nm are used, because it decreases the dynamic range available for the detection of the intended analytes. Higher sensitivity can be achieved with the photomultiplier design by scaling the photomultiplier voltage, and therefore, for measurements below 240 nm the Beckman optics are preferred. On the other hand, serial wavelength detection imposes significant throughput limitations, especially when more than 20 wavelengths or more than two samples need to be measured in a single run. Since the Cölfen optics permit the simultaneous acquisition of a broad range of closely-spaced wavelengths for multiple cells, these optics are eminently well suited for measuring systems in which chromophores need to be examined over a large wavelength range, especially in the visible range where the Cölfen optics have sufficient dynamic range. When using UltraScan to acquire multiwavelength data from the Beckman optics [19], data acquisition is restricted to a maximum of 100 wavelengths per cell, but they do not have to be spaced in regular intervals. However, 100 wavelengths are often too many, especially for rapidly sedimenting analytes, since significant delays are encountered during the initial calibration of the photomultiplier gain setting, which needs to be performed for each wavelength and channel. This delay prevents data collection at the beginning of the experiment, causing potential loss of detection for large molecules and aggregates. Consequently, the scan frequency for each wavelength is significantly decreased, despite rapid radial scanning. For experiments with more than 15–20 wavelengths, it is often not advisable to scan more than a single cell, while in the Cölfen optics, all rotor positions can be filled, and scans early in the experiment are not missed. With the Beckman optics, it may be necessary to sacrifice sedimentation resolution by scanning rapidly sedimenting analytes at slower than optimal rotor speed to gain more time for scanning. Nevertheless, the signal-to-noise level in the Beckman optics is exceptional, typically resulting in residual mean square deviations of less than 2.0×10^{-3} absorbance units, with a radial resolution of 0.001 cm. Hence, comparable statistics can be achieved with the Optima AUC even with fewer scans in a sedimentation velocity experiment.

2.5 Methods

2.5.1 Design of MW-AUC experiments

We describe here how the features of each optical system are best exploited for multi-wavelength analytical ultracentrifugation experiments involving biopolymers, in particular with a focus on macromolecular interactions. We focus on the experimental design and describe how the spectral features of each analyte can be used to optimize the information obtained. We also discuss the algorithms used to analyze multi-wavelength data obtained from the Optima AUC since they differ from the earlier described procedure that is suitable only for the Cölfen optics [6].

In addition to the information available from traditional single-wavelength methods, MW-AUC also characterizes the hydrodynamically separated molecules based on their spectral contributions, identifying free and complexed species they may form, as well as the stoichiometries of their complexes. This technique relies on the ability to spectrally separate the absorbing species present in a mixture. In order to successfully separate the spectral contributions from different analytes, several requirements need to be met. First, the mixing event should not induce a change of the analyte’s absorbance properties. For example, in the case of complex formation, the absorbance spectra of the interacting analytes should not red- or blue-shift or change molar absorptivity. Second, the absorbance spectra of the pure analytes should be known, preferably in molar dimensions, such that molar stoichiometries can be derived from the analysis. Third, the absorbance spectra of the analytes need to be sufficiently orthogonal in order to be linearly separable. This requirement can be checked by calculating the angle θ between the molar extinction vectors *u* and v of two analytes to be spectrally separated, given by (Equation 2.1):

$$\theta = cos^{-1}\left(\frac{u \cdot v}{\|u\| \cdot \|v\|}\right) \text{ where } u = \begin{bmatrix} \varepsilon_{u1} \\ \varepsilon_{u2} \\ \cdots \\ \varepsilon_{u3} \end{bmatrix} \text{ and } v = \begin{bmatrix} \varepsilon_{v1} \\ \varepsilon_{v2} \\ \cdots \\ \varepsilon_{v3} \end{bmatrix}$$

(Equation 2.1)

and ε_{ui} is the molar extinction coefficient of *u* at wavelength *i* and ε_{vi} the molar extinction coefficient of analyte *v* at the same wavelength. It is necessary that *u* and *v* contain the same wavelengths.

Theoretically, if the angle θ is larger than zero, the spectra can be separated. An angle of 90° indicates perfect orthogonality (for example, two spectra with baseline separation between their absorbance peaks), but angles much smaller than 90° can be separated. The degree of success depends on the total signal available and the quality of the data. In general, the larger the angle θ, the better the chance the analytes can be spectrally separated. For analytes where the absorbance spectra show significant overlap (small θ), it is often helpful to expand the measured wavelength range. For example, when comparing the absorbance spectra of a protein and DNA molecule, using just the typical 260 nm/ 280 nm absorbance pairs, θ is 27.8°, however, when considering the absorbance range between 230 and 300 nm with 2 nm increments (36 wavelengths), the angle increases to 42.7°, offering significantly improved resolution (also see

SF. 2.1). The final requirement is that molar extinction profiles are within the same order of magnitude, ensuring that the observed signal is comparable between the different species. This can be a challenge when the molar extinction coefficient of a protein at 280 nm is much less than the molar extinction coefficient of an interacting nucleic acid at 260 nm. In such cases, mixtures quickly reach the dynamic range of the detector without providing sufficient signal from the protein. A solution is to shift or expand the measured wavelength range.

For example, nucleic acids have a particularly strong extinction in the 250–260 nm region. This absorbance band partially overlaps with the 280 nm absorbance band of aromatic amino acids. Hence, measuring wavelength range between 240 and 300 nm is well suited to characterizing protein-nucleic acid interactions when proteins contain a large mole fraction of tryptophan and tyrosine, and the nucleic acids are short [9]. Systems with longer fragments of nucleic acids in a mixture with proteins containing a small mole fraction of tryptophan and tyrosine provide limited absorbance signal for multi-wavelength experiments conducted in this wavelength range, because the relatively small molar absorbance from aromatic amino acids is overwhelmed by the absorbance from the nucleic acid, and equimolar protein concentrations will be difficult to detect. In such cases, sufficient signal from the protein can be achieved by including wavelengths in the region between 215 and 240 nm, where the peptide bond absorbance provides significantly higher absorbance (see SF. 2.1). This equalizes the absorptivity between protein and nucleic acid and at the same time increases the orthogonality between the absorption profiles of protein and nucleic acid.

In all cases, it is important to use a buffer system that does not absorb significantly in the measured wavelength range. Suitable buffer systems include phosphate- or low concentration optically pure TRIS-based buffers, and those that do not contain absorbing additives such as nucleotides, chelators, or reductants.

For the Beckman optics, it is beneficial to minimize the number of wavelengths scanned because each wavelength has to be measured sequentially. That reduces the number of scans available for each individual wavelength compared to the Cölfen optics, which scans all wavelengths simultaneously in each radial scan. One approach to maximizing the orthogonality of the measured spectra, while minimizing the number of measured wavelengths, is to interpolate spectral regions in the absorbance spectrum that exhibit linear change and to measure only wavelengths required for a faithful interpolation of the spectrum. For example, in regions where the change in the spectrum is linear over multiple wavelengths, only the endpoints of the linear region need to be measured [14]. This will reduce the number of measured wavelengths and the time required to complete the scan cycle, thereby increasing the total number of scans collected for each wavelength. Another trick for the Beckman optics is to choose a rotor speed that is optimally synchronized with the flash lamp timing, which decreases the elapsed time between successive scans. The timing delays between scans, as a function of rotor speed, are calculated in the UltraScan data acquisition module for the Optima AUC [19] (see SF. 2.3), optimal rotor speeds include 14,600–14,900, 31,500–32,900, 45,800–50,900, and 59,600–60,000 RPM. In these ranges, scan times are 8 s/channel or less. Unfortunately, it is not possible in the Optima to scan only one channel of a cell. Therefore, for multi-wavelength AUC experiments acquired with the Optima AUC, it is advisable to run a single cell containing two samples, one in each channel sector, because a reference channel is not required when using UltraScan. Importantly, experiments should always be measured in intensity mode to reduce stochastic noise contributions to the data [12].

2.5.2 Identification of basis spectra

For hetero-associating systems, AUC can separate free and complexed species based on differences in their hydrodynamic properties. In addition to the hydrodynamic separation, optical deconvolution can identify the molar contribution of each interacting partner in a molecular complex, and provide the stoichiometry of binding [6, 9]. Reliable interpretation of the stoichiometry requires that accurate molar extinction coefficients are known for each analyte in the mixture contributing to the absorbance of the sample over the entire spectral range examined in a MW-AUC experiment. To obtain these molar extinction coefficient profiles, high-quality absorbance scans of each analyte are required. Depending on the spectral properties of the analyte, the dynamic range of the detector (0.1–0.9 OD) can be readily exceeded at some of the selected wavelengths when only a single analyte concentration is measured. For example, the molar extinction coefficient of a protein at 215 nm can easily exceed the extinction coefficient at 280 nm by 1–2 orders of magnitude when aromatic side chains are sparse or absent in the protein sequence (e.g., histones, collagen). To address this challenge, multiple dilutions need to be measured in the spectrophotometer. This approach ensures that overlapping wavelength ranges for one or more dilutions fall within the dynamic range of the detector, yielding a reliable intrinsic extinction profile over the entire wavelength range. To obtain the intrinsic extinction spectrum of an analyte over the entire wavelength range, the extinction profile fitter in UltraScan [5] is used to globally fit multiple dilution spectra from the analyte to sums of Gaussian terms using the Levenberg-Marquardt non-linear least squares fitting algorithm [20, 21] (see SF. 2.4). The fitted model is normalized with a known molar extinction coefficient (typically at 280 nm for proteins), which can be retrieved directly from the UltraScan LIMS database and derived from the associated protein sequence based on the molar absorptivity of the amino acid composition at 280 nm. The global molar extinction profile is used downstream to decompose experimental MW-AUC data into molar concentrations of spectral constituents (discussed below).

If the buffer used to dissolve the analytes absorbs in the measured wavelength range, then all absorbance measurements of the analytes of interest should be performed in a spectrophotometer blanked against the buffer. Also, since all MW-AUC experiments should be performed in intensity mode, the absorbing buffer must be considered as a separate spectral species in the downstream MW decomposition. In order to obtain its absorbance spectrum, the buffer's absorbance profile must be measured by blanking the spectrophotometer with distilled water. We recommend to use spectrophotometers with a 1 cm pathlength, fitted with quartz cuvettes. For all studies reported here we used a benchtop GENESYS™ 10S UV–Vis spectrophotometer from ThermoFisher.

For reversibly interacting systems, the thermodynamic binding isotherms are most reliably determined by measuring MW-AUC experiments of multiple titration points with different ratios of the interacting partners mixed together [9]. The spectral decomposition module in UltraScan is used to obtain the mixing ratio from each titration point. The absorbance spectrum of the titration mixture and the intrinsic molar extinction spectra for each distinct analyte in the mixture (the basis spectra) are loaded into the program. The program will determine the overlapping wavelengths available from each spectrum, and use this range to calculate the molar composition, providing residuals to the fit. The program also reports on the angle θ ((Equation 2.1) between each pair of basis spectra (see SF. 2.5). By monitoring θ, the program can also be used to optimize the wavelength selection to aid in the experimental design. If a hypochromic or hyperchromic shift occurs in the absorbance profile due to mixing, the fitting residuals will appear to be non-random, providing feedback on the suitability of including selected wavelength ranges in the decomposition.

2.5.3 Analysis of MW-AUC experiments

Due to the design differences between the two multi-wavelength optical systems, experimental data differ in their structure and need to be analyzed with different strategies. The Cölfen optics collect all wavelengths simultaneously and provide a complete spectrum for the entire wavelength range, which is determined by the diffraction grating used in the optics [4]. As a result, each radial observation in the scan simultaneously produces a complete wavelength scan, and all wavelengths for that radial point are collected at the same time. This produces a 3D scan image (absorbance as a function of wavelength and radius, see Fig. 2.1, left panel). This image can immediately be decomposed to obtain isolated optical signals for each separated analyte in the mixture [6] for each radial point in each scan. In the Beckman optics, data at different wavelengths are collected sequentially, which causes radial scans at each wavelength to be collected at a different time. The time difference observed between the first and last wavelength collected for a multi-wavelength scan depends on the rotor speed and the total number of wavelengths collected, and is estimated by UltraScan. The difference in time between individual scans at different wavelengths is not obviously apparent from visual inspection of the 3D data (see Fig. 2.1, right panel), but must be addressed before spectral decomposition can be performed.

For both optics, the analysis procedure before spectral decomposition is identical. The analysis starts by removing all systematic noise from each triple (a triple is a complete experimental dataset from a unique cell, channel, and wavelength) and fitting the boundary conditions (meniscus and bottom of the cell). At this point, sedimentation velocity data from each triple are processed separately. The analysis proceeds through several refinement steps. In the first refinement step, a two-dimensional spectrum analysis (2DSA) [22] is performed with simultaneous time-invariant noise subtraction. In the Optima AUC, intensity data obtained from a photomultiplier tube contains significant time-invariant noise contributions, which must be removed first. This intensity variation is less of an issue with the linear CCD array used in the Cölfen optics, but the same step is still recommended to remove time invariant noise resulting from other sources, such as imperfections in the optical path or scratches in the cell windows. In the next step, a second refinement is performed with the 2DSA, adding time- and radially-invariant noise correction, and fitting of the boundary conditions (meniscus and bottom of cell). Computing times for a 2DSA analysis varies based on grid resolution, data size, number of wavelengths to be analyzed, analysis sub-type, and compute resources available. A typical multi-wavelength scan has only 40–60 scans, since all acquired scans are distributed over all wavelengths. This significantly reduces the compute time to 10–30 s/triple on a modern multi-core architecture, such as a 128-core AMD Epyc server. Since UltraScan is highly parallelized, and implemented on NSF-XSEDE/Access high-performance computing infrastructure, all triples can often be analyzed in parallel, where each 2DSA analysis utilizes 16 cores. Hence, the more wavelengths are acquired, the fewer scans are collected per wavelength, and since all triples can be analyzed at the same time, the faster the individual analysis of a triple proceeds. The overall compute time is typically reduced for all wavelengths to less than 30 s for the entire calculation. Evaluation of boundary conditions (meniscus, bottom radius) requires typically one order of magnitude more time, but only a single triple from each dataset needs to be analyzed, since all wavelengths from the same channel have identical boundary conditions. Typically, a wavelength with optimal signal-to-noise (absorbance between 0.5 and 0.7 OD, and RMSD values less than

0.002) is chosen to determine the meniscus and bottom position. 2DSA-IT calculations are on the same order of magnitude, and need to be performed for each triple. Deconvolution is a very fast process, typically finishing within 10–20 s. On account of the mirror optics, both optical systems are essentially free of chromatic aberration [7]. However, in the Optima AUC, chromatic aberration in some instruments is large enough to require correction. This is handled in the UltraScan software by uploading a chromatic aberration profile into the LIMS database, which is applied to all data acquired from the Optima AUC during the data import stage. This process is further discussed by Stoutjesdyk et al. [23]. After chromatic aberration correction, the boundary condition fitting step only needs to be performed on a single wavelength from each channel, and the fitted positions are applied to the edit profiles of all other wavelengths in that dataset, since they are all from the same channel and therefore have the same boundary conditions. For best results in the boundary condition fit, it is recommended to select a wavelength which contains sufficient signal and low stochastic noise contributions. In the final refinement, an iterative 2DSA is performed with simultaneous time- and radially-invariant noise correction for each triple [22]. The simultaneous processing of hundreds of triples for multiple channels and wavelengths is best performed on a supercomputer, where all triples can be analyzed in parallel and in batch mode [24]. Sedimentation and frictional ratio parameter ranges for the 2DSA fits should be carefully adjusted to capture all hydrodynamic species in the sample. Fits for all triples should be inspected to ensure the fits result in random residuals for the entire AUC experiment by overlaying raw data and fitted models using the Finite Element Model Viewer in UltraScan. Root mean square deviations (RMSD) values for good fits range between 0.0015 and 0.003 absorbance units on the Optima AUC, and 0.0025–0.006 on the Cölfen optics, depending on wavelength and total absorbance at that wavelength. RMSD values are listed in the UltraScan

"Finite Element Model Viewer" and the analysis report for each 2DSA-IT fit. Residuals are displayed in the same module, as well as a red-green residual bitmap to aid in the assessment of fit quality. At this point, all systematic noise contributions should be removed from the data, and the final 2DSA refinement can be expected to be an accurate representation of the underlying data. The analytes contained in the 2DSA models will faithfully reproduce the hydrodynamic profiles from the experimental data, and the random residuals in the fitted data should only represent the stochastic noise contributions and have Gaussian distributions.

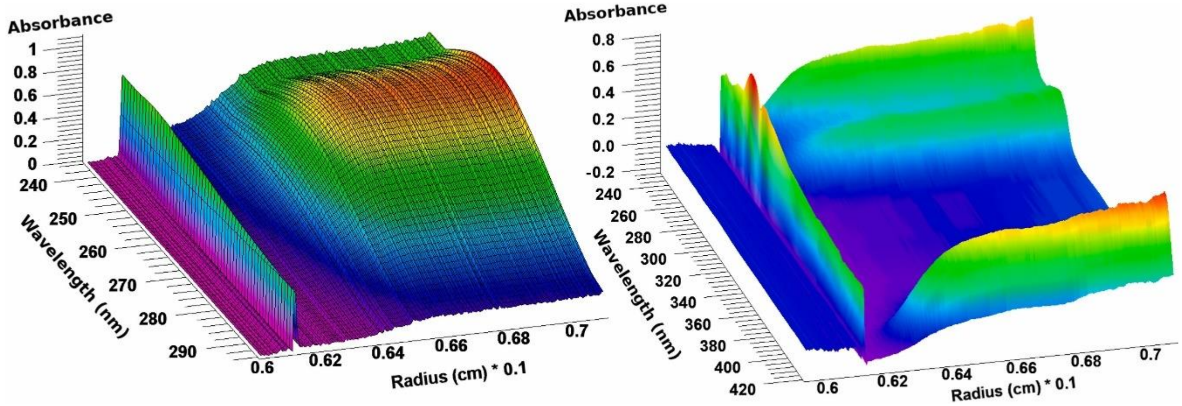


Figure 2.1: Multi-wavelength AUC data from a protein-RNA mixture acquired in the Cölfen optics (left) and a heme protein acquired from the Beckman optics (right).

2.5.4 Generation of a synchronous time grid for Optima AUC data

Before multi-wavelength data are decomposed into spectral basis vectors, one additional step is required with data from the Beckman optics. All wavelength data from the same channel must be transposed onto a synchronous time grid to handle the time discrepancies incurred during sequential wavelength acquisition. This is accomplished by loading the iterative 2DSA models from each triple belonging to a single channel into the Optima multi-wavelength fit simulator (started from the "Multiwavelength" menu in UltraScan's main menu). Using the fitted, iterative 2DSA models, this module simulates the entire MW-AUC experiment, such that all triples from different wavelengths are now on a common and synchronous time grid. The synchronous time grid ensures that each scan from every wavelength has the same time stamp and can be used to obtain a reliable wavelength scan for each radial position. During the simulation of the synchronous time grid, the user can set all specifics of this simulation (rotor speed, meniscus position, run duration, number of scans) to match the settings of the original experiment (further described in SF. 2.6). Partial concentrations of all analytes will be faithfully reproduced from the 2DSA models. Next, the simulations are uploaded to the LIMS database and edited to produce an equivalent MW-AUC experiment to the original experimental dataset. There is no requirement to add stochastic or other systematic noise to the data since all noise components have already been identified and subtracted from the data in earlier refinement steps. At this point, the data from the Cölfen optics and the Beckman optics are equivalent and can be further processed by the spectral decomposition module.

2.5.5 Spectral decomposition of MW-AUC data

Spectral decomposition of MW-AUC data resolves species with unique chromophores in a mixture based on their individual absorbance properties. Data processed as described above result in a set of three dimensional surfaces for each scan's time point in the experiment, representing the absorbance as a function of wavelength and radius. Hence, for each surface, each radial position of this surface gives rise to a wavelength scan. This wavelength scan can be decomposed into its basis absorbance spectra as described earlier and shown in SF. 2.5. The decomposition is accomplished by using the non-negatively constrained least squares algorithm (NNLS) developed by Lawson and Hansen [25]. It assures that only positive contributions, or zero, are generated during the decomposition. For each basis vector, a two-dimensional (2D) space-time sedimentation velocity dataset will be generated during this process. Together, all basis vectors solve the linear equation subject to the constraint $x_i > 0$ (Equation 2.2):

$$A_j = x_1 v_1 + x_2 v_2 + \cdots x_n v_n$$

(Equation 2.2)

where A_j is the absorbance wavelength scan at data point j, composed of spectral vectors v_i with amplitudes x_i. After processing all data in a MW-AUC dataset, the decomposition results in n traditional 2D sedimentation velocity experiments, each representing a separate, unique spectral species in the mixture. If the spectral vectors contain molar extinction coefficients, the resulting 2D datasets inherit units of molar concentration. If only relative concentrations are needed, a normalized absorbance scale can be used as well. The decomposition is carried out by the UltraScan "MWL Species Fit" module from the "Multiwavelength" menu in the UltraScan main menu. This process is further detailed in SF 2.7. The resulting traditional 2D datasets (molar concentration as a function of radius and time) for each spectral component can be uploaded to the UltraScan LIMS system, edited, and analyzed by standard UltraScan procedures (2DSA [22], PCSA [26], GA [18, 27], van Holde-Weischet [28] or other methods available in UltraScan). There is no further need to fit the boundary conditions, remove systematic noise contributions, or perform a Monte Carlo noise analysis. Comparing spectrally separated hydrodynamic analyses will reveal both free and complexed species, where species with identical hydrodynamic parameters represent complexes. Importantly, when integrating each spectral species found in a complex, the molar stoichiometry of the species in that complex is revealed, as long as the spectral basis vectors are expressed in terms of molar extinction coefficients [9].

2.5.6 Preparations and experimental design of oilseed protein extracts

Oilseed material was extracted by homogenizing with an aqueous solution. The liquid phase was separated from the fibers using a centrifuge decanter. The extract was defatted using a disc-stack separator and the resulting protein dispersion underwent pH modification and centrifugal clarification to remove insoluble aggregates. The remaining soluble protein was recovered by concentrating via ultrafiltration and diluted with phosphate buffered saline to approximately 1.0 OD at 280 nm for analytical ultracentrifugation. The sample was placed into a 2-channel 1.2 cm epon charcoal centerpiece fitted with quartz windows (Beckman Coulter, Indianapolis) and measured in an AN50Ti rotor at position 4, opposite a counterbalance, and spun for 13 h at 50,000 rpm, and at 20 °C. 34 wavelengths were collected for each sample between 250 and 349 nm, with 3 nm increments.

2.5.7 Preparations and experimental design of fluorescent proteins

Hexahistidine-tagged fluorescent proteins (mPapaya, Teal, and Ultramarine) were expressed in *E. coli* and purified by Ni-Sepharose chromatography according to methods described earlier [29-31]. All proteins were buffer exchanged to standard phosphate buffered saline (PBS) for analytical ultracentrifugation measurements. The absorbance of the three protein samples was adjusted to 0.5 OD at each protein's peak absorbance wavelength in the visible region (see SF. 2.2). Samples were then mixed at the following ratios: TFP:UFP (5:1, 1:1, 1:3); UFP: TFP:mPapaya (1:3:3); UFP:TFP:mPapaya (1:1:1). Up to two samples were measured simultaneously in each run by scanning one cell. The samples were placed into the two channels of a 2-channel 1.2 cm epon charcoal centerpiece fitted with quartz windows (Beckman Coulter, Indianapolis) and measured in an AN50Ti rotor at position 4, opposite a counterbalance, and spun for 12 h at 46,000 rpm at 20 °C. Ninety-one wavelengths were collected for each sample between 420 and 600 nm, with 2 nm increments.

Fluorescent proteins and oilseed samples were measured at the Canadian Center for Hydrodynamics at the University of Lethbridge with an Optima AUC instrument. Data analysis was performed on the Chinook cluster at the University of Lethbridge.

2.6 Results

The hydrodynamic separation of free and associated analytes alone often does not provide sufficient resolution to permit a clear and unambiguous interpretation of AUC experiments for two important reasons: First, different analytes may have similar hydrodynamic properties, such as size, degree of globularity, and density, and therefore would not be distinguishable by hydrodynamic separation alone. Secondly, the ability to uniquely identify each analyte decreases with an increasing number of analytes because the observed signal is proportional to the relative amount of each analyte. If too many analytes are present, it is impossible to distinguish them based on hydrodynamic information alone. In MW-AUC experiments, the additional spectral information provides a second dimension to identify analytes by their unique chromophores. We distinguish two basic experiments: a) cases where spectral properties are not available in pure form for each unique chemical species present in a mixture, and b) cases where the pure spectra for each unique chemical species are known, and molar extinction coefficients are available for each measured wavelength. In the case of (a), it is still possible to extract and review the spectral properties after hydrodynamically separating all species. Even though molar extinction coefficients may not be available, the spectral pattern may still provide useful insights. For cases described by (b), a mathematical deconvolution of spectral contributors will then identify the chemical nature of each hydrodynamic species, determine the relative amount of each species, and for complexes, the stoichiometry of assembly. Representative examples for both cases are presented below. In a MW-AUC experiment, multiple datasets from traditional single-wavelength experiments are collected at multiple wavelengths and combined for a global analysis, which characterizes the analytes in a second, spectral dimension, based on their spectral contributions to the overall signal. Since different types of biopolymers have unique spectral properties, it is therefore possible to resolve them not only based on their hydrodynamic properties, but also based on their absorbance spectra.

2.6.1 Hydrodynamic separation of spectral components

In cases where absorbance spectra from individual analytes with unique spectral characteristics are not available in pure form for all components in a mixture, an optical deconvolution of individual analytes will not be possible. Instead, a different strategy can prove valuable. A plot of the sedimentation coefficient distribution as a function of wavelength displays the spectral profiles of the hydrodynamically separated species and may allow identification of some or all species in the mixture by reviewing the shape of their absorbance spectra. This approach can be very effective and useful, provided multiple components in the mixture can be hydrodynamically separated. A representative example of this approach was demonstrated for mixtures of CdTe quantum dots by Karabudak et al. [32], where 24 unique hydrodynamic species were identified, and unique spectral properties of at least seven components could be derived over the examined wavelength range. In this method, s-values with non-zero amplitude obtained at different wavelengths are integrated at each wavelength to generate a spectral absorbance pattern for each unique hydrodynamic species.

The hydrodynamic separation of biopolymers typically has a lower size resolution than the highly dense metal quantum dots. However, if hydrodynamic separation is achieved, this method is still effective for classifying individual components. UltraScan offers a three-dimensional (3D) viewer, which projects the integrated sedimentation profile as a function of wavelength, or an integrated 2D plot of the absorbance profile at a particular sedimentation coefficient.

Example 1. – Identification of components in an oil seed protein extract. Fig. 2.2 shows analytical ultracentrifugation experiments performed with a heterogeneous oil seed plant protein extract after removing the lipid phase. In this example, the plant extract contains water-soluble polyphenols, small molecules with a 315 nm absorbance peak, as well as proteins of unknown size. When single wavelength experiments from two different wavelengths are compared (280 and 340nm), the hydrodynamic results are strikingly different, producing apparently contradictory results (Fig. 2.2, panel A). The absorbances of the 0.5–2.5S, and the 12.5S species differ greatly, despite having been measured from the same sample. The reason for this discrepancy is the presence of different chemicals which possess dissimilar chromophores, and therefore different molar extinction coefficients at the same wavelengths. As can be seen from panel A, the information derived even from a dual wavelength experiment is insufficient to provide an unambiguous explanation of the observed results. However, in addition to providing a hydrodynamic separation of the species present in this sample, a MW-AUC experiment successfully answers the following questions:

1. What are the identities of the species sedimenting at 0.5–12.5S?

2. Does the sample contain polyphenols, proteins, or both?

3. Are there interactions between polyphenols and the protein components?

4. Are the protein components degraded or intact?

First, the majority of the polyphenols, identified by their 315 nm absorbance peak, sedimented, as expected, with a very low sedimentation coefficient (~0.5–1.0S) and did not appear to be bound to any larger molecules. However, a small amount of protein degradation products was visible between 1.0 and 2.5S, with a spectral signature of a protein molecule, potentially also bound to a small amount of polyphenol. A peak around 12.5S displayed a spectral signature of a typical protein with an absorbance maximum around 280 nm. The protein sedimenting at 12.5S has a narrow size distribution, suggesting that this protein is intact, and the absence of 315 nm absorbance indicates that no polyphenols are bound. A smaller amount of protein signal was found at < 2S, suggesting the presence of a second protein or a small fraction of a potentially degraded protein. Its absorbance spectrum suggests that a small amount of polyphenols could be binding to the degraded protein samples.

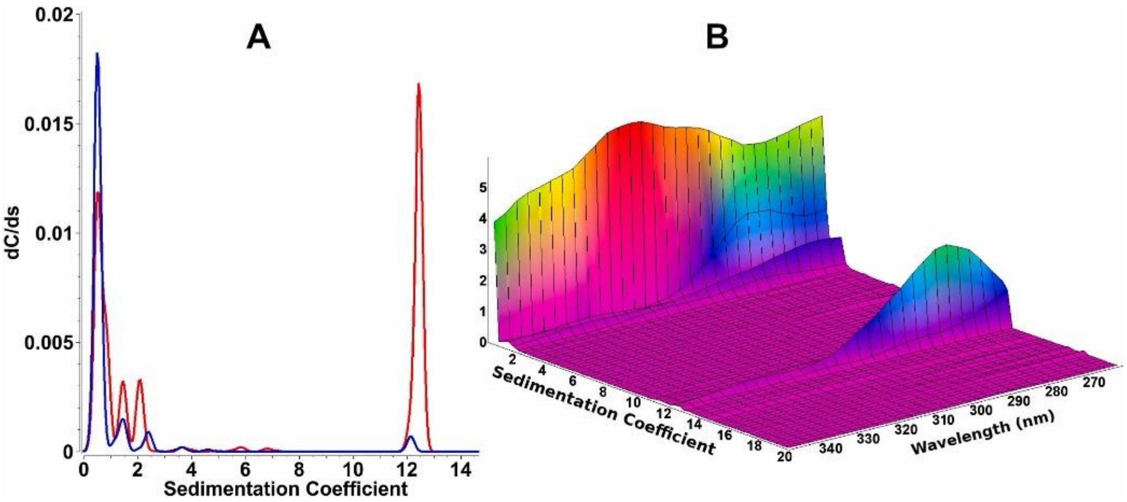


Figure 2.2: Sedimentation velocity experiment of oil seed protein extracts. Panel A: Traditional single-wavelength analytical ultracentrifugation analysis (using 2DSA-IT analysis) at 340 nm (blue trace) and 280nm (red trace). The two wavelengths clearly result in completely different concentrations for the hydrodynamic species present in the same sample, making an unambiguous interpretation impossible when considering a single, or only a few wavelengths. Panel B: MW-AUC analysis of the same experiment shown in panel A, covering 260–350 nm. A 3-dimensional representation of absorbance as a function of sedimentation coefficient as well as wavelength is much more informative (composite generated from 2DSA-IT models). For example, the 0.8 S species is easily identified as a polyphenol because of its 315 nm absorbance maximum, while the 12.5S species displays the typical absorbance profile of a protein, with an absorbance maximum at ~280 nm. Smaller species sedimenting at 1.5–2.2S appear to be degraded proteins, with some associated polyphenols. All data shown were collected with the Beckman optical system.

2.6.2 Spectral separation of hydrodynamic components

If pure spectra are available for individual species in a mixture, along with their molar extinction coefficients, spectral decomposition can be applied to determine absolute molar amounts of each species, whether free in solution or interacting with another molecule. In this case, the stoichiometry of an interaction is also available. A large class of experimental applications lend themselves to this approach.

Example 2. – Use of fluorescent tags: To study biopolymers without distinct chromophores (lipids, carbohydrates) or protein-protein interactions between proteins with very similar absorbance profiles in the ultraviolet, fluorescent tags or fluorescent protein fusions can be used to impart a unique chromophore to each interaction partner molecule. Excitation spectra from commercially available fluorescent dyes for tagging biopolymers and fluorescent proteins span a wide range of the visible spectrum, and they can be used to add a unique chromophore to any molecule of interest. By mixing multiple, differentially tagged molecules, it is possible to study the assembly of complex, multi-protein systems, and follow the order of assembly as well as the stoichiometry of assembly. To validate the performance of this approach, we mixed ultramarine [29], mTeal [30], and mPapaya [31] fluorescent proteins at different ratios and measured their sedimentation between 400 and 600 nm, the region containing the most significant difference in their absorbance spectra (see SI 2.2). After spectrally deconvoluting the MW-AUC experimental data, all three species can be baseline-resolved and accurately quantified (see Fig. 2.3, panels 1–4). In contrast, it was impossible to distinguish the components from a single wavelength analysis of an equimolar mixture of the same three molecules measured at a single wavelength of 486 nm(see Fig. 2.3, panel 5). When analyzed by MW-AUC, varying ratios of absorbance recovered from the peak integrations shown in Fig. 2.3 accurately reflect the pipetted ratios, confirming the ability of the method to resolve complex mixtures accurately. This result is even more remarkable considering that mTeal and mPapaya have identical hydrodynamic properties and therefore are not distinguishable when measured using traditional single-wavelength AUC. Unlike the monomeric mTeal and mPapaya, Ultramarine sediments as a constitutive dimer with a higher sedimentation coefficient, allowing it to be hydrodynamically separated.

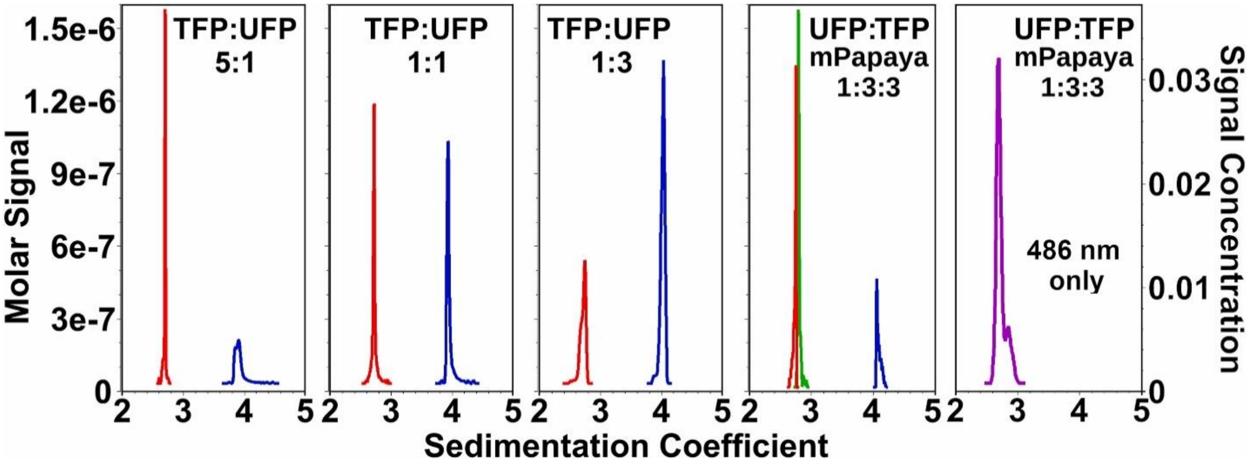


Figure 2.3: MW-AUC analysis resolves mixtures of fluorescent proteins. MW-AUC analysis of mixtures of two or three fluorescent proteins mixed at different ratios: Ultramarine fluorescent protein (UFP, blue), mTeal fluorescent protein (TFP, red), and mPapaya (green). Relative ratios of mixed proteins can be resolved within pipetting error. Panels 1–5 from left to right: 1: TFP:UFP 5:1, 2: TFP:UFP 1:1, 3: TFP:UFP 1:3, 4: UFP:TFP:mPapaya 1:3:3.5: UFP:TFP:mPapaya 1:3:3 (purple), measured at 486 nm only (see SF. 2.2 for spectral overlap at 486 nm). All proteins have identical monomeric mass, but UFP exists as a constitutive dimer, which results in a higher sedimentation coefficient. Note that loading concentrations of mPapaya and TFP are correctly resolved despite having identical sedimentation coefficients, but only when analyzed by MW-AUC. A single wavelength analysis of a 1:1:1 mixture of all three proteins only produces an indistinguishable single peak at 486 nm, where TFP and mPapaya have maximum overlap, because UFP barely absorbs at this wavelength, and TFP and mPapaya have the same sedimentation coefficient. All analyses shown here were generated from differential g(s) profiles derived from van Holde-Weischet analysis of the spectrally decomposed datasets, except the results in Panel 5, which are based on a traditional AUC experiment at 486 nm.

2.7 Discussion

The MW-AUC method extends the capabilities of an important biophysical characterization tool by adding a spectral characterization dimension to the hydrodynamic separation traditionally achieved by AUC. As is documented in examples representing two classes of MW-AUC experiments, distinct advantages are realized in the resolution and information content when studying heterogeneous systems where multiple analytes with unique chromophores are present in mixtures. MW-AUC adds a new tool to study and validate biopolymer interactions under physiological solution conditions, and to measure equilibrium constants for hetero-interactions by sedimentation velocity experiments [33]. As demonstrated in example 2 (see Fig. 2.3) highly precise molar quantities can be measured for each component in a complex mixture, which is required for accurate determination of assembly stoichiometry and thermodynamic interaction coefficients. Even if pure basis spectra of each analyte in a mixture are not available, this method can still prove useful in understanding the composition of an unknown sample, as is demonstrated in example 1 (see Fig. 2.2). Taken together, these new capabilities provide important new avenues for the solution-based investigation of complex, interacting systems by providing higher resolution details about composition, binding strength, and stoichiometry of interaction than can be achieved with traditional single wavelength or Rayleigh interference AUC approaches, or even with other solution-based methods. Our results point to the potential of significantly improved solution characterization possibilities, and open the door to novel investigation of complex systems, promising revolutionary insights. New instrumentation available in the form of the Cölfen and Beckman optical systems, as well as software advances in the UltraScan software contribute to the advances reported here, and provide convenient access to this technology. The open-source UltraScan software used for the analysis of MW-AUC experiments can be freely retrieved from Github in source code or binary format for Windows, Macintosh and Linux computers [34].

2.8 References

1. Braun P, Gingras A-C. History of protein–protein interactions: From egg-white to complex networks. Proteomics. 2012;12(10):1478-1498. doi: https://doi.org/10.1002/pmic.201100563.

2. Brautigam CA, Padrick SB, Schuck P. Multi-Signal Sedimentation Velocity Analysis with Mass Conservation for Determining the Stoichiometry of Protein Complexes. PLoS ONE. 2013;8(5):e62694. doi: 10.1371/journal.pone.0062694.

3. Strauss HM, Karabudak E, Bhattacharyya S, Kretzschmar A, Wohlleben W, Cölfen H. Performance of a fast fiber based UV/Vis multiwavelength detector for the analytical ultracentrifuge. Colloid and Polymer Science. 2008;286(2):121-128. doi: 10.1007/s00396-007-1815-5.

4. Pearson JZ, Krause F, Haffke D, Demeler B, Schilling K, Cölfen H. Chapter One - Next-Generation AUC Adds a Spectral Dimension: Development of Multiwavelength Detectors for the Analytical Ultracentrifuge. In: Cole JL, editor. Methods Enzymol. 562: Academic Press; 2015. p. 1-26.

5. Demeler B, Gorbet GE. Analytical ultracentrifugation data analysis with UltraScan-III. In Analytical Ultracentrifugation: Instrumentation, Software, and Applications; Uchiyama, S., Stafford, W. F., Laue, T., Eds.; . Analytical Ultracentrifugation: Springer; 2016. p. 119-143.

6. Gorbet GE, Pearson JZ, Demeler AK, Cölfen H, Demeler B. Next-generation AUC: analysis of multiwavelength analytical ultracentrifugation data. Methods Enzymol. 562: Elsevier; 2015. p. 27-47.

7. Pearson J, Hofstetter M, Dekorsy T, Totzeck M, Cölfen H. Design concepts in absorbance optical systems for analytical ultracentrifugation. Analyst. 2018;143(17):4040-4050. Epub 2018/07/06. doi: 10.1039/c8an00706c. PubMed PMID: 29975381.

8. Karabudak E, Wohlleben W, Cölfen H. Investigation of β-carotene–gelatin composite particles with a multiwavelength UV/vis detector for the analytical ultracentrifuge. Eur Biophys J. 2010;39(3):397-403. doi: 10.1007/s00249-009-0412-6.

9. Zhang J, Pearson JZ, Gorbet GE, Cölfen H, Germann MW, Brinton MA, et al. Spectral and hydrodynamic analysis of West Nile Virus RNA–protein interactions by multiwavelength sedimentation velocity in the analytical ultracentrifuge. Anal Chem. 2017;89(1):862-870.

10. Cölfen H, Wohlleben W, Walter J. A multi-wavelength detector for the Beckman Optima XL. OpenAUC Documents and License Terms. https://wiki.aucsolutions.com/openAUC/.

11. Walter J, Sherwood PJ, Lin W, Segets D, Stafford WF, Peukert W. Simultaneous analysis of hydrodynamic and optical properties using analytical ultracentrifugation equipped with multiwavelength detection. Anal Chem. 2015;87(6):3396-3403. Epub 2015/02/14. doi: 10.1021/ac504649c. PubMed PMID: 25679871.

12. Demeler B. Methods for the design and analysis of sedimentation velocity and sedimentation equilibrium experiments with proteins. Current protocols in protein science. 2010;60(1):7.13. 11-17.13. 24.

13. Johnson CN, Gorbet GE, Ramsower H, Urquidi J, Brancaleon L, Demeler B. Multi-wavelength analytical ultracentrifugation of human serum albumin complexed with porphyrin. Eur Biophys J. 2018;47(7):789-797.

14. Hu J, Hernandez Soraiz E, Johnson CN, Demeler B, Brancaleon L. Novel combinations of experimental and computational analysis tested on the binding of metalloprotoporphyrins to albumin. Int J Biol Macromol. 2019;134:445-457. doi: https://doi.org/10.1016/j.ijbiomac.2019.05.060.

15. Koebke KJ, Kühl T, Lojou E, Demeler B, Schoepp-Cothenet B, Iranzo O, et al. The pH-Induced Selectivity Between Cysteine or Histidine Coordinated Heme in an Artificial α-Helical Metalloprotein. Angew Chem Int Ed Engl. 2021;60(8):3974-3978. Epub 2020/11/21. doi: 10.1002/anie.202012673. PubMed PMID: 33215801; PubMed Central PMCID: PMCPMC7902473.

16. Salveson PJ, Haerianardakani S, Thuy-Boun A, Yoo S, Kreutzer AG, Demeler B, et al. Repurposing Triphenylmethane Dyes to Bind to Trimers Derived from Aβ. J Am Chem Soc. 2018;140(37):11745-11754. doi: 10.1021/jacs.8b06568.

17. Horne CR, Henrickson A, Demeler B, Dobson RC. Multi-wavelength analytical ultracentrifugation as a tool to characterise protein–DNA interactions in solution. Eur Biophys J. 2020;49(8):819-827.

18. Brookes E, Demeler B. Parsimonious Regularization Using Genetic Algorithms Applied to the Analysis of Analytical Ultracentrifugation Experiments. GECCO ’07: Proceedings of the 9th Annual Conference on Genetic and Evolutionary Computation, vols. 361–368, July 2007. ACM 978-1-59593-697-4/07/0007. 2007. doi: https://doi.org/10.1145/1276958.1277035.

19. Savelyev A, Gorbet GE, Henrickson A, Demeler B. Moving analytical ultracentrifugation software to a good manufacturing practices (GMP) environment. PLoS Comput Biol. 2020;16(6):e1007942. doi: 10.1371/journal.pcbi.1007942.

20. Levenberg K. A method for the solution of certain non-linear problems in least squares. Q Appl Math. 1944;2(2):164-168. doi: https://doi.org/10.1090/qam/10666.

21. Marquardt DW. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. Journal of the Society for Industrial and Applied Mathematics. 1963;11(2):431-441. doi: 10.1137/0111030. PubMed PMID: doi:10.1137/0111030.

22. Brookes E, Cao W, Demeler B. A two-dimensional spectrum analysis for sedimentation velocity experiments of mixtures with heterogeneity in molecular weight and shape. Eur Biophys J. 2010;39(3):404-414.

23. Stoutjesdyk M, Henrickson A, Minors G, Demeler B. A calibration disk for the correction of radial errors from chromatic aberration and rotor stretch in the Optima AUC™ analytical ultracentrifuge. Eur Biophys J. 2020. doi: 10.1007/s00249-020-01434-z.

24. Brookes E, Demeler B. Parallel computational techniques for the analysis of sedimentation velocity experiments in UltraScan. Colloid and Polymer Science. 2008;286:139-148.

25. Lawson CL, Hanson RJ. Solving least squares problems: SIAM; 1995.

26. Gorbet G, Devlin T, Hernandez Uribe Blanca I, Demeler Aysha K, Lindsey Zachary L, Ganji S, et al. A Parametrically Constrained Optimization Method for Fitting Sedimentation Velocity Experiments. Biophys J. 2014;106(8):1741-1750. doi: https://doi.org/10.1016/j.bpj.2014.02.022.

27. Brookes E, Demeler B. Genetic algorithm optimization for obtaining accurate molecular weight distributions from sedimentation velocity experiments. Analytical Ultracentrifugation VIII, Progr Colloid Polym Sci Springer. 2006;131:78-82.

28. Demeler B, Van Holde KE. Sedimentation velocity analysis of highly heterogeneous systems. Anal Biochem. 2004;335(2):279-288.

29. Li Y, Forbrich A, Wu J, Shao P, Campbell RE, Zemp R. Engineering Dark Chromoprotein Reporters for Photoacoustic Microscopy and FRET Imaging. Scientific Reports. 2016;6(1):22129. doi: 10.1038/srep22129.

30. Ai H-w, Henderson JN, Remington SJ, Campbell Robert E. Directed evolution of a monomeric, bright and photostable version of Clavularia cyan fluorescent protein: structural characterization and applications in fluorescence imaging. Biochem J. 2006;400(3):531-540. doi: 10.1042/bj20060874.

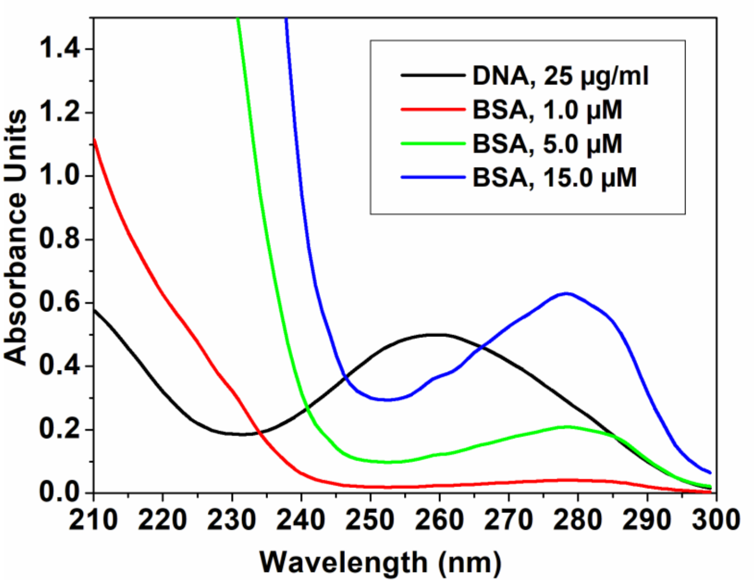
31. Hoi H, Howe Elizabeth S, Ding Y, Zhang W, Baird Michelle A, Sell Brittney R, et al. An Engineered Monomeric Zoanthus sp. Yellow Fluorescent Protein. Chemistry & Biology. 2013;20(10):1296-1304. doi: https://doi.org/10.1016/j.chembiol.2013.08.008.

32. Karabudak E, Brookes E, Lesnyak V, Gaponik N, Eychmüller A, Walter J, et al. Simultaneous Identification of Spectral Properties and Sizes of Multiple Particles in Solution with Subnanometer Resolution. Angew Chem Int Ed Engl. 2016;55(39):11770-11774. Epub 2016/07/28. doi: 10.1002/anie.201603844. PubMed PMID: 27461742; PubMed Central PMCID: PMCPMC5148131.

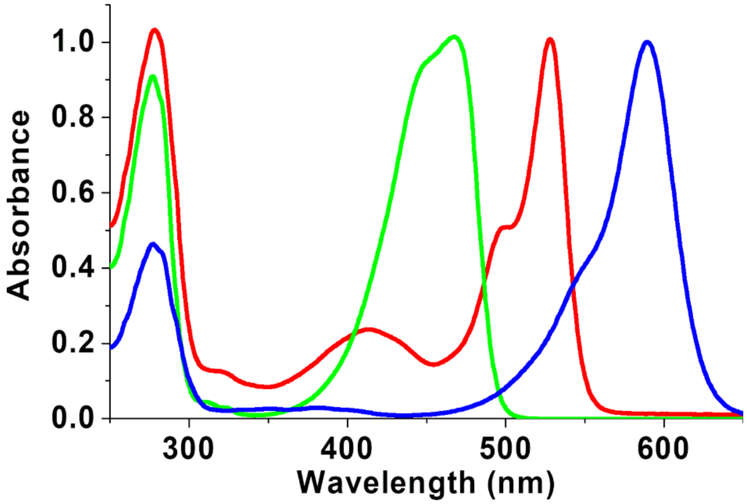
33. Demeler B, Brookes E, Wang R, Schirf V, Kim CA. Characterization of reversible associations by sedimentation velocity with UltraScan. Macromol Biosci. 2010;10(7):775-782. Epub 2010/05/21. doi: 10.1002/mabi.200900481. PubMed PMID: 20486142.

34. UltraScan GitHub archive for free, open source analysis software for MW-AUC experiments. https://github.com/ehb54/ultrascan3.

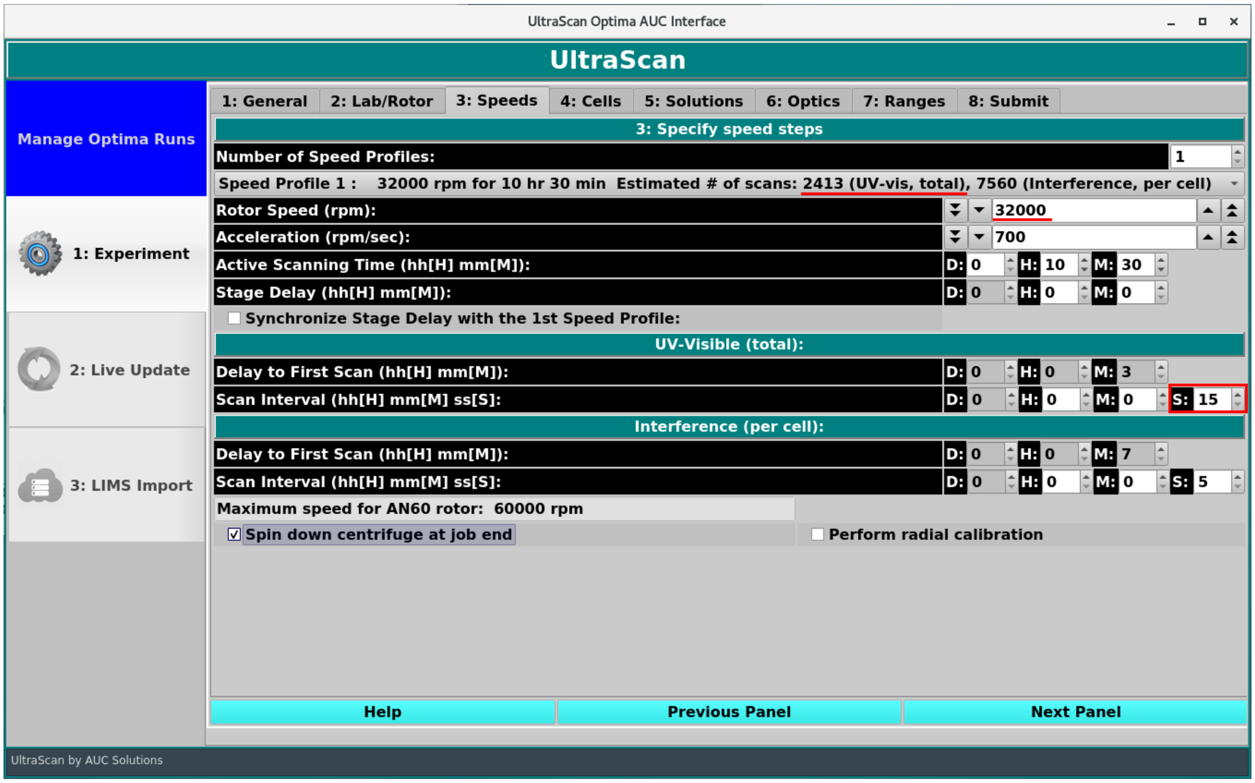
2.9 Supplemental Information



SF 2.1: Absorption spectra for DNA and protein. Absorbance profiles for 25 μg/ml DNA (black) and protein (Bovine serum albumin (BSA), blue: high concentration (15.0 μM), green: medium concentration (5.0 uM), red: low concentration (1.0 μM) between 210-300 nm. Even low concentration proteins, or proteins without aromatic side chains, provide sufficient signal and spectral orthogonality when wavelengths between 210-240 nm are included due to the absorbance from the peptide bond in the protein's backbone. BSA concentrations above 5 μM exceed the dynamic range of the detector below 235 nm.



SF 2.2: Absorbance Spectra for Fluorescent Protein Stocks. Ultramarine (blue), mTFP1 (green), and mPapaya (red). While all fluorescent proteins share a peak at 280 nm due to tryptophan and tyrosine absorbance, the fluorescence excitation spectra in the visible region are markedly different and can be used to easily distinguish the spectra in a multi-wavelength AUC experiment. Proteins can be expressed as fusion proteins with fluorescent proteins to inherit unique excitation spectra from fluorescent proteins.



3: Specify speed steps

Number of Speed Profiles: 1

Speed Profile 1 : 32000 rpm for 10 hr 30 min Estimated # of scans: 2413 (UV-vis, total), 7560 (Interference, per cell)

Rotor Speed (rpm): 32000

Acceleration (rpm/sec): 700

Active Scanning Time (hh[H] mm[M]): D: 0 H: 10 M: 30

Stage Delay (hh[H] mm[M]): D: 0 H: 0 M: 0

☐ Synchronize Stage Delay with the 1st Speed Profile:

UV-Visible (total):

Delay to First Scan (hh[H] mm[M]): D: 0 H: 0 M: 3

Scan Interval (hh[H] mm[M] ss[S]: D: 0 H: 0 M: 0 S: 15

Interference (per cell):

Delay to First Scan (hh[H] mm[M]): D: 0 H: 0 M: 7

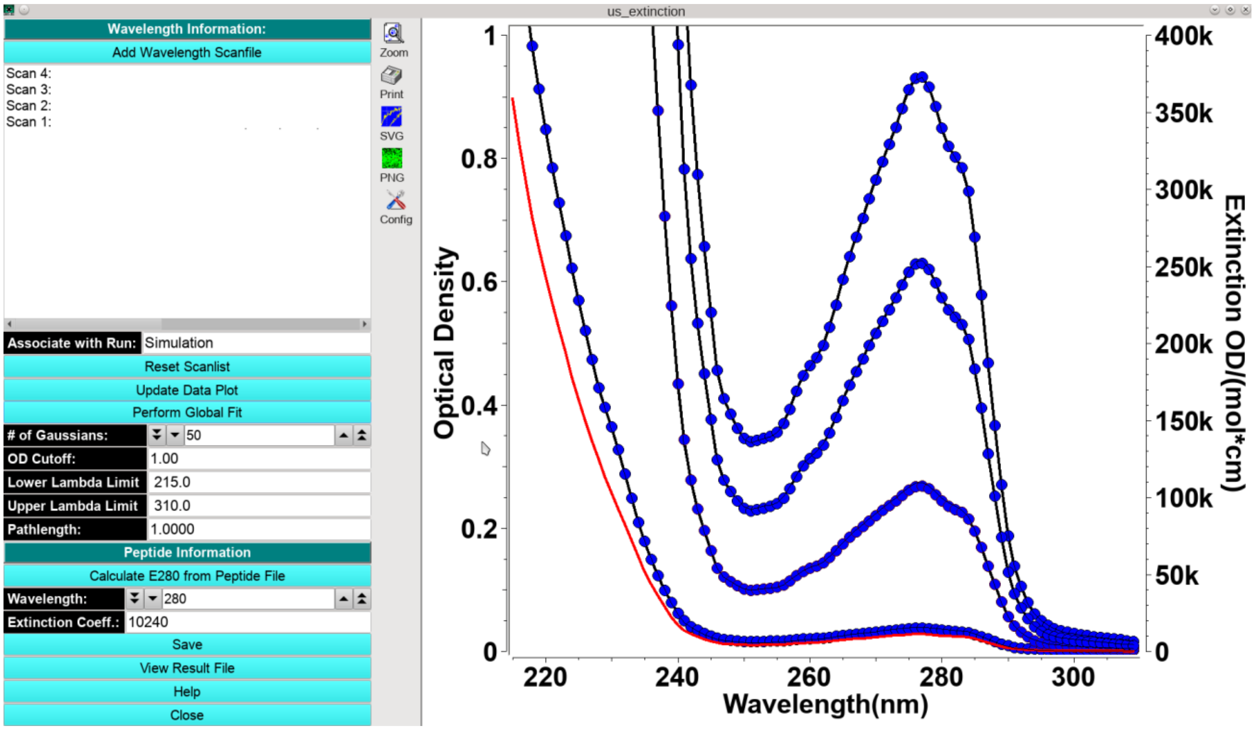
Scan Interval (hh[H] mm[M] ss[S]: D: 0 H: 0 M: 0 S: 5

Maximum speed for AN60 rotor: 60000 rpm

☑ Spin down centrifuge at job end

☐ Perform radial calibration

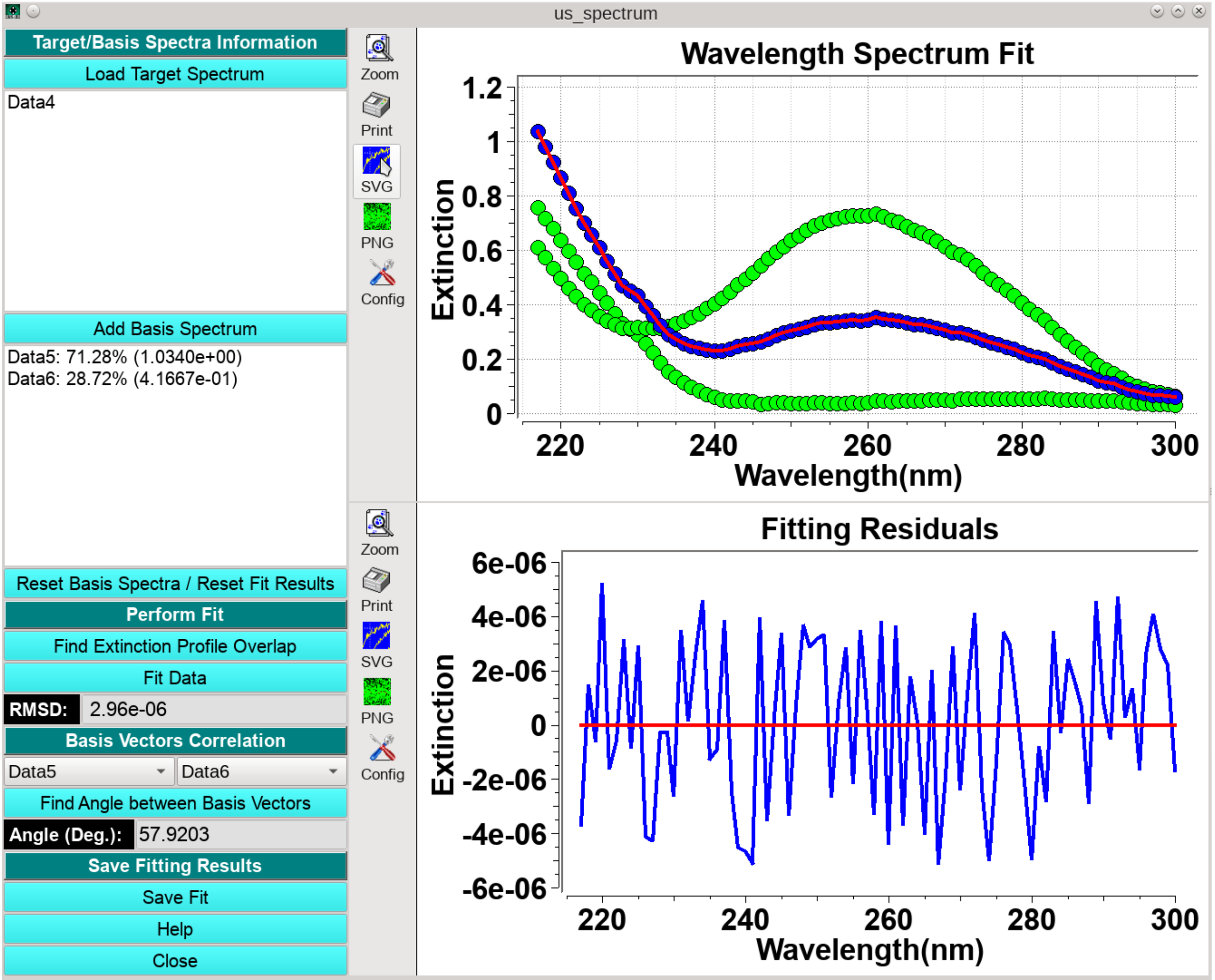
SF 2.3: Speed selection control in the UltraScan data acquisition interface for the Optima AUC. Items marked in red relate the rotor speed selected for an experiment to the time interval in seconds between sequential scans, and the total number of scans that can be acquired over the user-selected data acquisition run time.



Scan 4:
Scan 3:
Scan 2:
Scan 1:

Associate with Run: Simulation
# of Gaussians: 50
OD Cutoff: 1.00
Lower Lambda Limit 215.0
Upper Lambda Limit 310.0
Pathlength: 1.0000
Wavelength: 280
Extinction Coeff.: 10240

SF 2.4: Global spectrum fitting interface. Experimental absorbance wavelength data from multiple analyte dilutions (blue dots, left y-axis) are fitted to a global molar extinction profile (red line, right y-axis). The global molar extinction profile model is precisely scaled to each concentration that was measured in a spectrophotometer (black lines). The global molar extinction profile is scaled using the known molar extinction coefficient at a single wavelength. This value can be obtained from the literature, or for proteins, can be estimated from the protein sequence based on the cumulative molar absorptivities of all tryptophan, tyrosine, phenylalanine, and cysteine residues contained in the protein. Two or more globally fitted molar extinction profiles are used for the decomposition of a multi-wavelength AUC experiment. To check the orthogonality of two spectra, the angle between the two profiles can be calculated by using the two molar extinction profiles as the vectors u and v in Equ 2.1.



Data4

Data5: 71.28% (1.0340e+00)
Data6: 28.72% (4.1667e-01)

RMSD: 2.96e-06

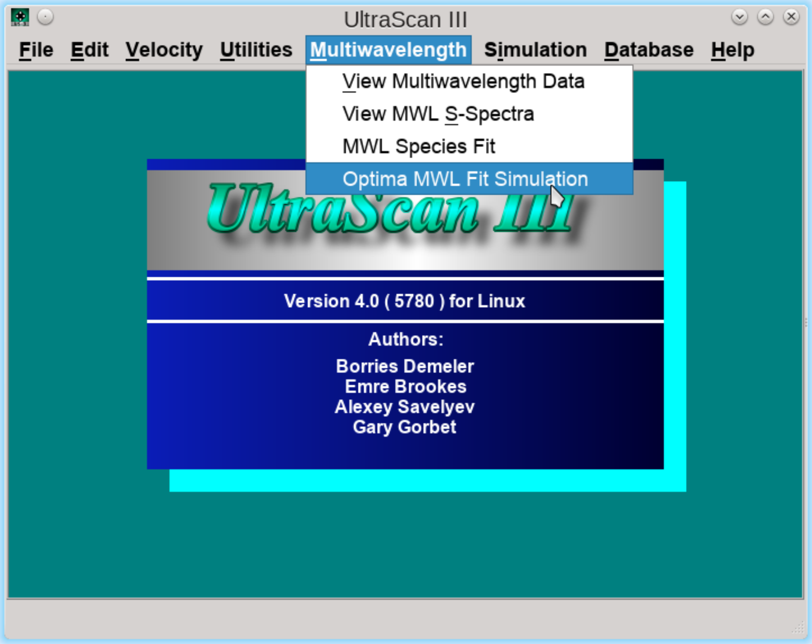
Data5 Data6

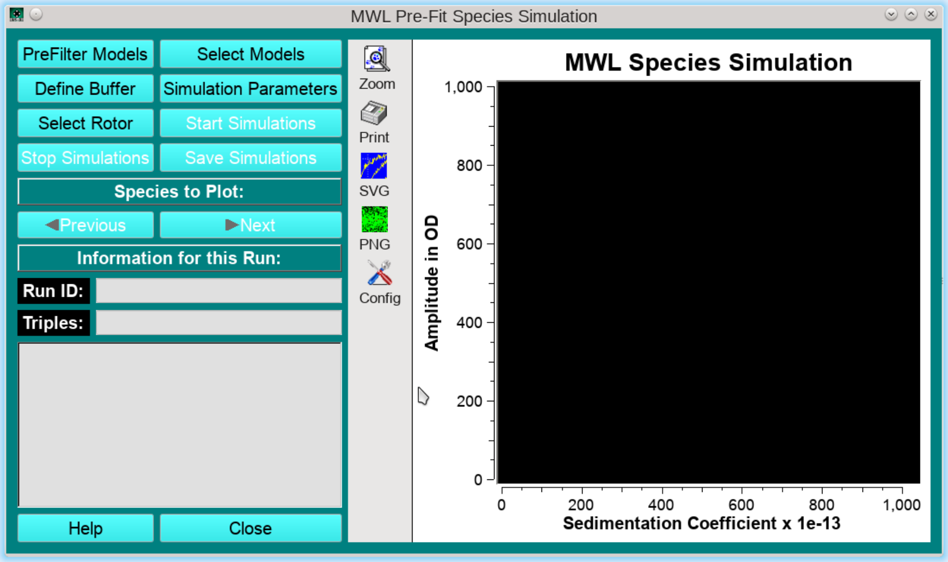
Angle (Deg.): 57.9203

SF 2.5: Spectrum decomposition utility in UltraScan. Absorbance scans of mixtures of spectrally diverse analytes (blue dots) are decomposed into their spectral basis vectors (green dots) and fitted to a linear combination (red line, above. Residuals of the fit are shown in the lower panel (blue line). Relative contribution of each basis is computed and displayed in the left panel. The angle between two basis vectors is displayed in the lower left.

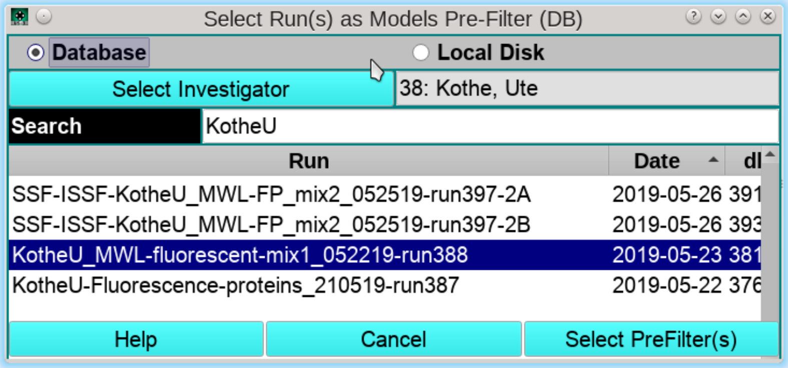
SF 2.6: Step-by-Step instructions for the generation of time-synchronous multi-wavelength data from Optima AUC (Beckman Optics) intensity data.

Step 1: Open the “Optima MWL Fit Simulation module from the main UltraScan menu:





Step 2: Select a prefilter for the 2DSA-IT models to be simulated by clicking on "*PreFilter Models*", and select the desired MW-AUC Optima AUC experiment:



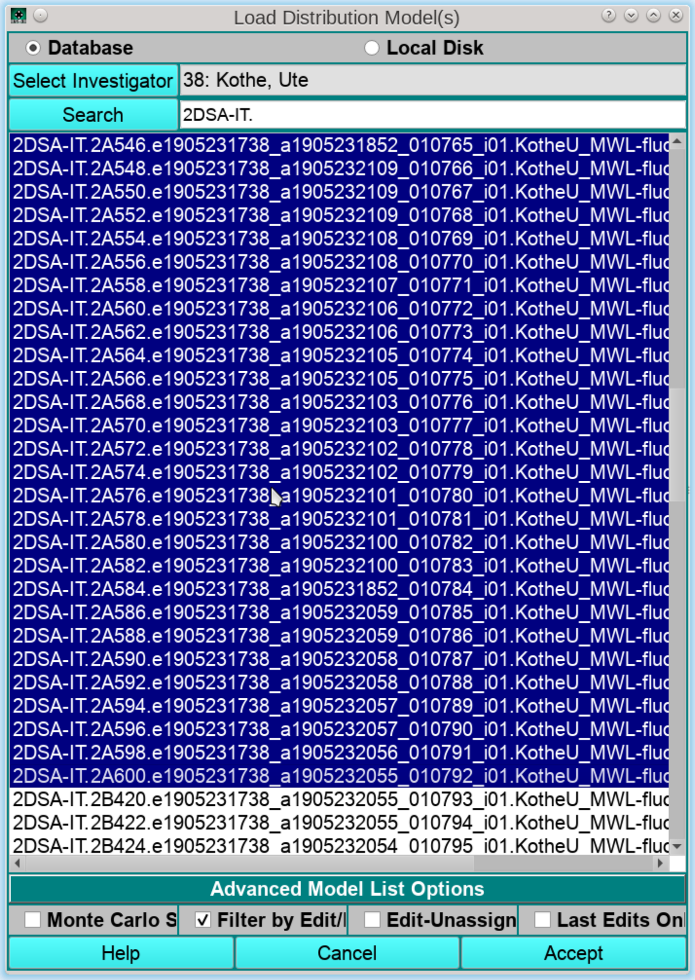
● Database ○ Local Disk

Select Investigator: 38: Kothe, Ute

Search: KotheU

Run	Date	dl
SSF-ISSF-KotheU_MWL-FP_mix2_052519-run397-2A	2019-05-26	391
SSF-ISSF-KotheU_MWL-FP_mix2_052519-run397-2B	2019-05-26	393
KotheU_MWL-fluorescent-mix1_052219-run388	2019-05-23	381
KotheU-Fluorescence-proteins_210519-run387	2019-05-22	376

Step 3: Select all 2DSA-IT models for all wavelengths belonging to a single channel. By default, the program will display only 2DSA-IT models:



● Database ○ Local Disk

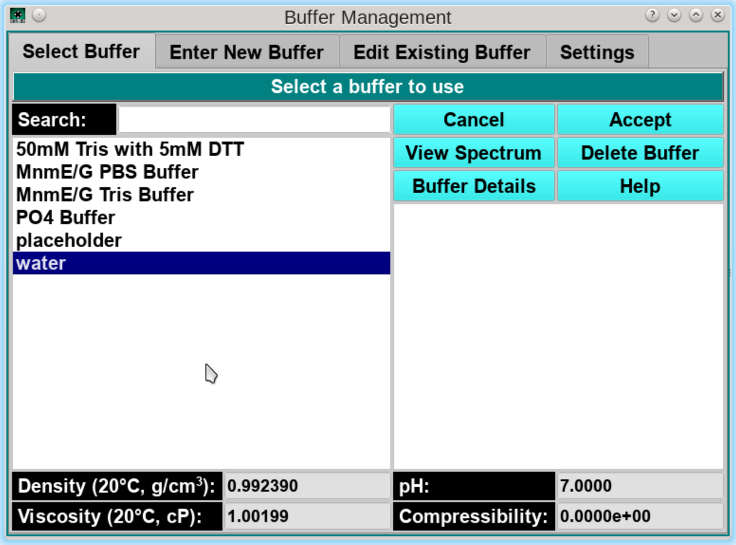
Select Investigator: 38: Kothe, Ute

Search: 2DSA-IT.

```
2DSA-IT.2A546.e1905231738_a1905231852_010765_i01.KotheU_MWL-fluc
2DSA-IT.2A548.e1905231738_a1905232109_010766_i01.KotheU_MWL-fluc
2DSA-IT.2A550.e1905231738_a1905232109_010767_i01.KotheU_MWL-fluc
2DSA-IT.2A552.e1905231738_a1905232109_010768_i01.KotheU_MWL-fluc
2DSA-IT.2A554.e1905231738_a1905232108_010769_i01.KotheU_MWL-fluc
2DSA-IT.2A556.e1905231738_a1905232108_010770_i01.KotheU_MWL-fluc
2DSA-IT.2A558.e1905231738_a1905232107_010771_i01.KotheU_MWL-fluc
2DSA-IT.2A560.e1905231738_a1905232106_010772_i01.KotheU_MWL-fluc
2DSA-IT.2A562.e1905231738_a1905232106_010773_i01.KotheU_MWL-fluc
2DSA-IT.2A564.e1905231738_a1905232105_010774_i01.KotheU_MWL-fluc
2DSA-IT.2A566.e1905231738_a1905232105_010775_i01.KotheU_MWL-fluc
2DSA-IT.2A568.e1905231738_a1905232103_010776_i01.KotheU_MWL-fluc
2DSA-IT.2A570.e1905231738_a1905232103_010777_i01.KotheU_MWL-fluc
2DSA-IT.2A572.e1905231738_a1905232102_010778_i01.KotheU_MWL-fluc
2DSA-IT.2A574.e1905231738_a1905232102_010779_i01.KotheU_MWL-fluc
2DSA-IT.2A576.e1905231738_a1905232101_010780_i01.KotheU_MWL-fluc
2DSA-IT.2A578.e1905231738_a1905232101_010781_i01.KotheU_MWL-fluc
2DSA-IT.2A580.e1905231738_a1905232100_010782_i01.KotheU_MWL-fluc
2DSA-IT.2A582.e1905231738_a1905232100_010783_i01.KotheU_MWL-fluc
2DSA-IT.2A584.e1905231738_a1905231852_010784_i01.KotheU_MWL-fluc
2DSA-IT.2A586.e1905231738_a1905232059_010785_i01.KotheU_MWL-fluc
2DSA-IT.2A588.e1905231738_a1905232059_010786_i01.KotheU_MWL-fluc
2DSA-IT.2A590.e1905231738_a1905232058_010787_i01.KotheU_MWL-fluc
2DSA-IT.2A592.e1905231738_a1905232058_010788_i01.KotheU_MWL-fluc
2DSA-IT.2A594.e1905231738_a1905232057_010789_i01.KotheU_MWL-fluc
2DSA-IT.2A596.e1905231738_a1905232057_010790_i01.KotheU_MWL-fluc
2DSA-IT.2A598.e1905231738_a1905232056_010791_i01.KotheU_MWL-fluc
2DSA-IT.2A600.e1905231738_a1905232055_010792_i01.KotheU_MWL-fluc
2DSA-IT.2B420.e1905231738_a1905232055_010793_i01.KotheU_MWL-fluc
2DSA-IT.2B422.e1905231738_a1905232055_010794_i01.KotheU_MWL-fluc
2DSA-IT.2B424.e1905231738_a1905232054_010795_i01.KotheU_MWL-fluc
```

☐ Monte Carlo S ☑ Filter by Edit/I ☐ Edit-Unassign ☐ Last Edits Onl

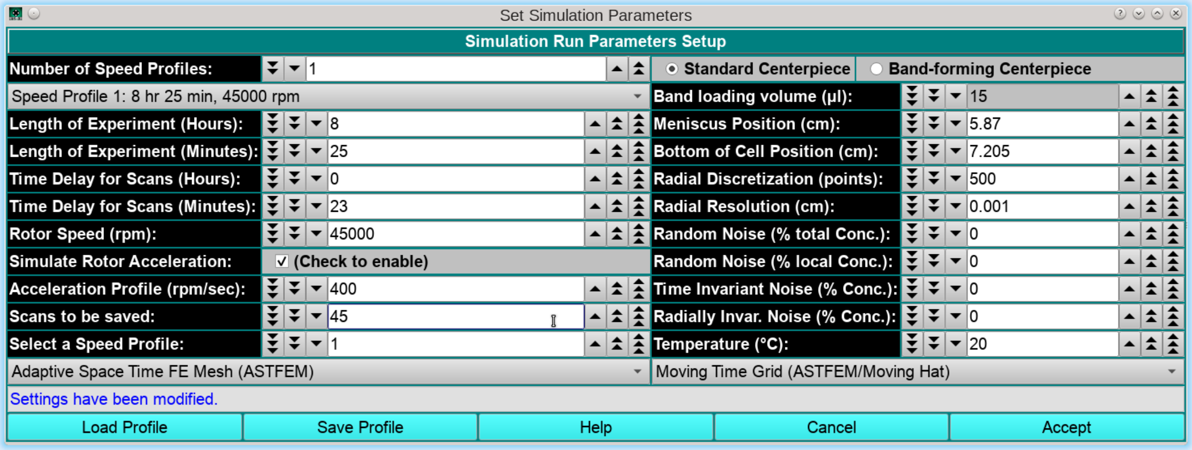
Step 4: Define a buffer by clicking on "*Define Buffer*". Since buffer density and viscosity were already taken into account during the original 2DSA analysis, all 2DSA-IT models are already corrected for standard conditions, i.e., water at 20ºC. Therefore, the user can pick water at 20ºC for all subsequent analysis steps as a buffer:



50mM Tris with 5mM DTT
MnmE/G PBS Buffer
MnmE/G Tris Buffer
PO4 Buffer
placeholder
water

Density (20°C, g/cm^3):	0.992390	pH:	7.0000
Viscosity (20°C, cP):	1.00199	Compressibility:	0.0000e+00

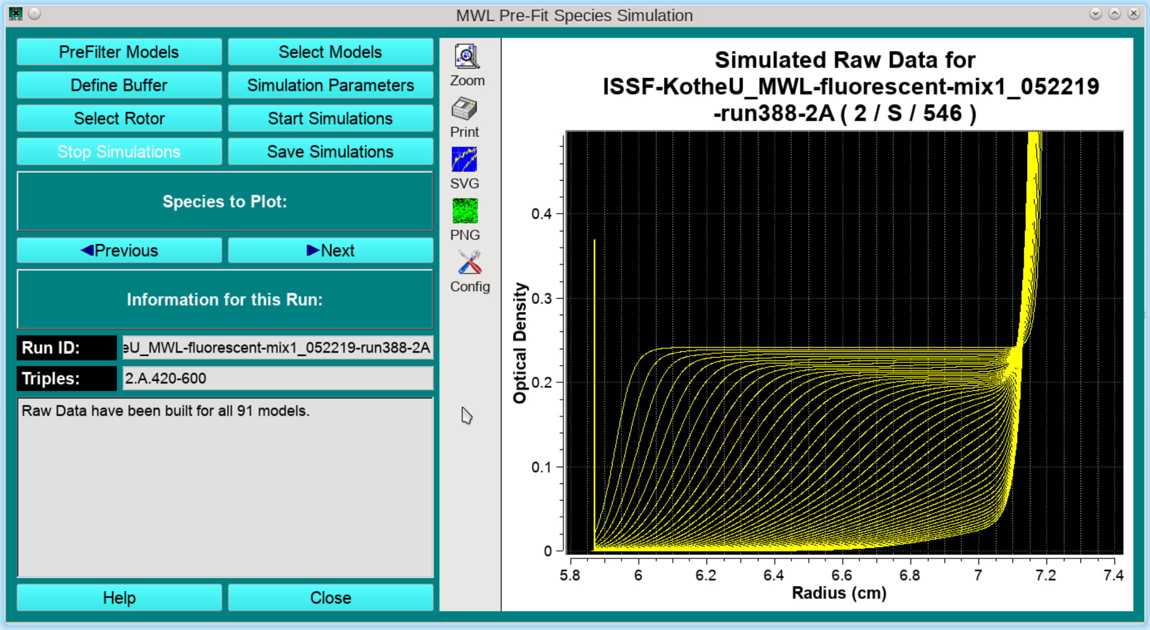
Step 5: In this step, the "*Simulation Parameters*" need to be defined. Ideally, these parameters should be identical to the experiment's run parameters, including rotor speed, meniscus position (can be retrieved from the associated edit profile after the meniscus fit), rotor type and calibration profile, number of scans, run duration and scan delay should all be adjusted. It is important to note that UltraScan will report sedimentation and diffusion coefficients already corrected to standard conditions (20ºC and water), so any simulations using the previously fitted 2DSA-IT models should use standard conditions:



Number of Speed Profiles:	1	● Standard Centerpiece	○ Band-forming Centerpiece
Speed Profile 1: 8 hr 25 min, 45000 rpm		Band loading volume (µl):	15
Length of Experiment (Hours):	8	Meniscus Position (cm):	5.87
Length of Experiment (Minutes):	25	Bottom of Cell Position (cm):	7.205
Time Delay for Scans (Hours):	0	Radial Discretization (points):	500
Time Delay for Scans (Minutes):	23	Radial Resolution (cm):	0.001
Rotor Speed (rpm):	45000	Random Noise (% total Conc.):	0
Simulate Rotor Acceleration:	☑ (Check to enable)	Random Noise (% local Conc.):	0
Acceleration Profile (rpm/sec):	400	Time Invariant Noise (% Conc.):	0
Scans to be saved:	45	Radially Invar. Noise (% Conc.):	0
Select a Speed Profile:	1	Temperature (°C):	20
Adaptive Space Time FE Mesh (ASTFEM)		Moving Time Grid (ASTFEM/Moving Hat)	

Settings have been modified.

Step 6: Perform the simulation by clicking on "*Start Simulation*". The program will re-simulate the fitted 2DSA-IT models with the experimental parameters defined in the simulation settings, generating a separate sedimentation velocity experiment for each wavelength. Now, each scan from each wavelength will be simulated for precisely the same time in all datasets. The simulated datasets will be shown in the graph windows, and a synthetic meniscus is generated as well to aid in downstream editing of these data. Clicking on "*Previous*" or "*Next*" allows the user to review each simulated dataset:



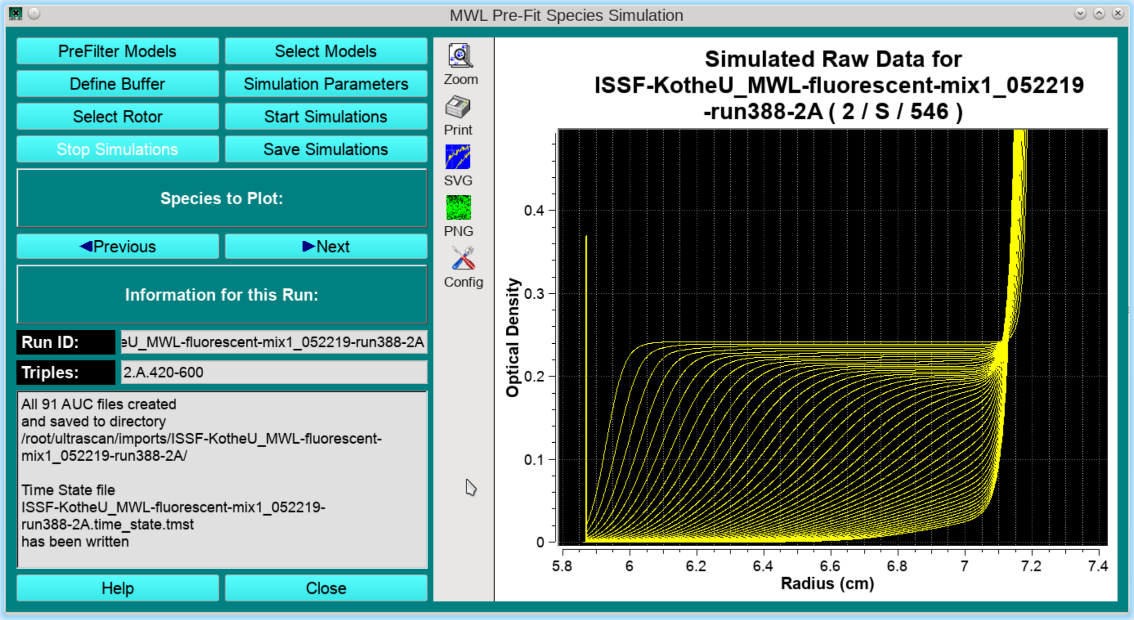
Run ID: eU_MWL-fluorescent-mix1_052219-run388-2A

Triples: 2.A.420-600

Raw Data have been built for all 91 models.

Simulated Raw Data for ISSF-KotheU_MWL-fluorescent-mix1_052219 -run388-2A (2 / S / 546)

Step 7: Saving the data. Once the simulated data have been reviewed, the data can be saved by clicking on "*Save Simulations*". Data will be written to the $HOME/ultrascan/imports folder into a subfolder that starts with prefix "*ISSF-*" (=initial simulated scan files):



Run ID: eU_MWL-fluorescent-mix1_052219-run388-2A

Triples: 2.A.420-600

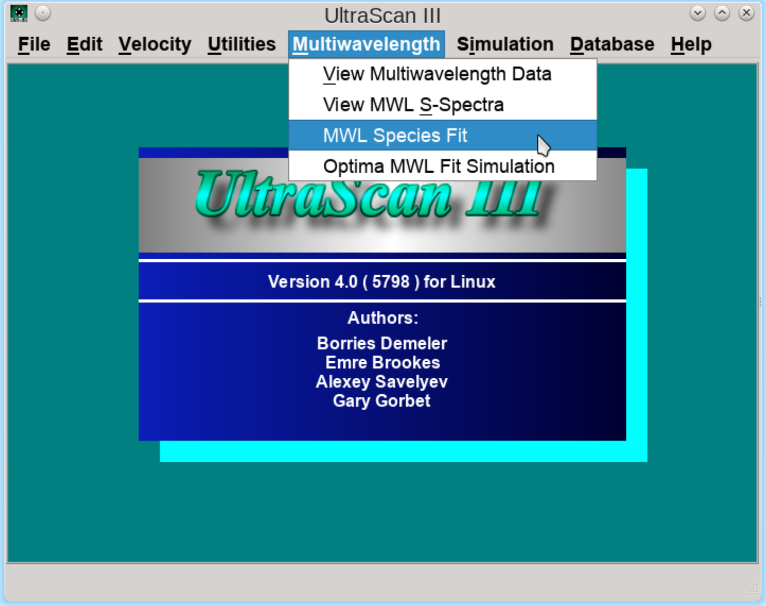
```
All 91 AUC files created
and saved to directory
/root/ultrascan/imports/ISSF-KotheU_MWL-fluorescent-
mix1_052219-run388-2A/

Time State file
ISSF-KotheU_MWL-fluorescent-mix1_052219-
run388-2A.time_state.tmst
has been written
```

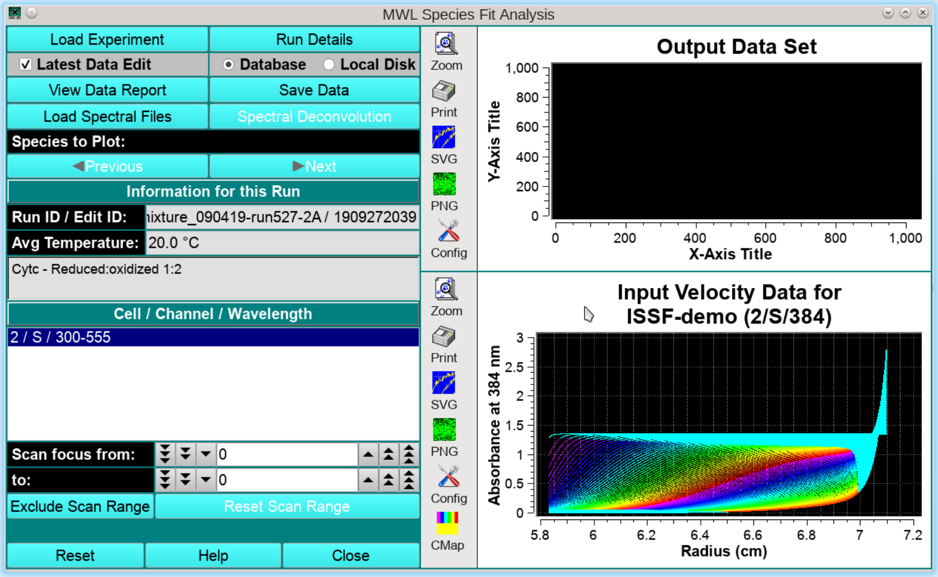
At this point, the ISSF data should be imported and edited like an ordinary MW-AUC velocity dataset.

SF 2.7: Step-by-Step instructions for the decomposition of time-synchronous multi-wavelength data

Decomposition of multi-wavelength sedimentation velocity experiments requires two or more extinction profiles for spectrally unique analytes present in a mixture, as well as a time-synchronous multi-wavelength dataset, either from the Cölfen optics or an ISSF dataset obtained after processing as described in SI-6, from the Beckman optics. The decomposition program is loaded from the main "*Multiwavelength*" menu entry by selecting "*MWL Species Fit*":



In the first step, the Cölfen optics or ISSF data are loaded into the program by clicking on "*Load Experiment*":



Run ID / Edit ID: ixture_090419-run527-2A / 1909272039

Avg Temperature: 20.0 °C

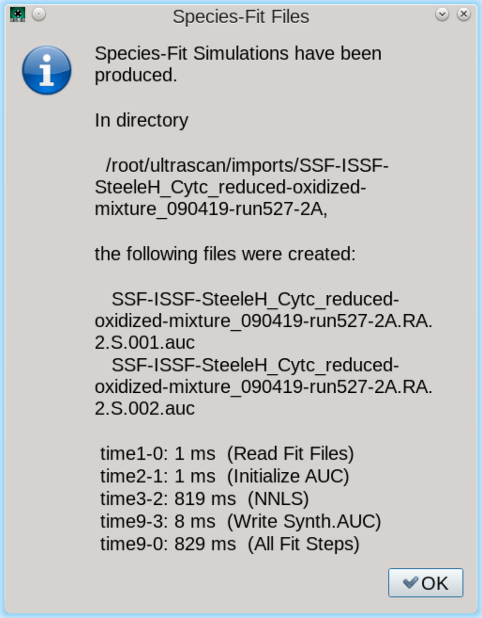
Cytc - Reduced:oxidized 1:2

2 / S / 300-555

Scan focus from: 0

to: 0

In the next step, the spectral basis vectors are loaded by clicking on "*Load Spectral Files*". A minimum of two spectral basis vectors need to be loaded. Once they are loaded, the "Spectral Deconvolution" button becomes active and should be clicked to start the deconvolution into separate datasets. The progress is reported in a dialog:



Species-Fit Simulations have been produced.

In directory

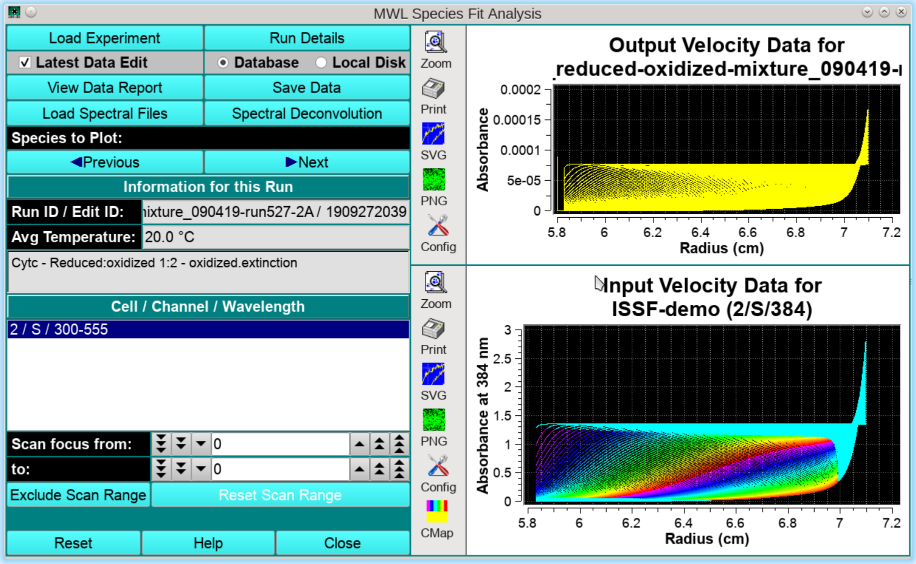
/root/ultrascan/imports/SSF-ISSF-SteeleH_Cytc_reduced-oxidized-mixture_090419-run527-2A,

the following files were created:

SSF-ISSF-SteeleH_Cytc_reduced-oxidized-mixture_090419-run527-2A.RA.2.S.001.auc
SSF-ISSF-SteeleH_Cytc_reduced-oxidized-mixture_090419-run527-2A.RA.2.S.002.auc

```
time1-0: 1 ms  (Read Fit Files)
time2-1: 1 ms  (Initialize AUC)
time3-2: 819 ms  (NNLS)
time9-3: 8 ms  (Write Synth.AUC)
time9-0: 829 ms  (All Fit Steps)
```

Clicking on "*OK*" will reveal the deconvoluted datasets in the upper panel and activates the "*Previous*" and "*Next*" buttons to switch between datasets, and activates "*Save Data*" to save the results:



Run ID / Edit ID: ixture_090419-run527-2A / 1909272039
Avg Temperature: 20.0 °C
Cytc - Reduced:oxidized 1:2 - oxidized.extinction
Cell / Channel / Wavelength: 2 / S / 300-555
Scan focus from: 0
to: 0

The saved data need to be imported into the UltraScan LIMS server, edited and then they can be analyzed without any further meniscus or noise processing.

Chapter 3. Protein – DNA interactions characterized by multi-wavelength AUC

Ishtiyaq Ahmed[1], Jeanette Hahn[2], Amy Henrickson[3], Faisal Tarique Khaja[4], Borries Demeler [3,5], David Dubnau [1,2], and Matthew B. Neiditch[1]

[1] Department of Microbiology, Biochemistry, and Molecular Genetics, New Jersey Medical School, Rutgers Biomedical Health Sciences, Newark, NJ 07103, USA.

[2] Public Health Research Institute, Rutgers Biomedical Health Sciences, Newark, NJ 07103, USA.

[3] Department of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, AB T1K 3M4, Canada.

[4] Greehey Children's Cancer Research Institute, University of Texas Health at San Antonio, San Antonio, TX 78229, USA.

[5] Department of Chemistry and Biochemistry, University of Montana, Missoula, MT 59801, USA.

3.1 Contributions of Authors

This chapter is adapted from a manuscript published in *Nature Communications*, volume 13, article number 7724 (2022), https://doi.org/10.1038/s41467-022-35129-0. The work was performed in collaboration with Ishtiyaq Ahmed, Dr. Jeanette Hahn, Dr. Faisal Tarique Khaja, Dr. David Dubnau, and Dr. Matthew B. Neiditch from Rutgers University, who developed the project and performed x-ray crystallography, EMSA, and *in vivo* experiments. Dr. Demeler and I conceived the AUC experiments and assisted with writing the AUC sections. I performed and analyzed the AUC data. In this chapter, I have copied the manuscript sections I helped write, with permission from the corresponding author, and rewrote the introduction and discussion sections.

3.2 Introduction

Horizontal gene transfer via transformation occurs in Gram-positive and Gram-negative bacteria [1]. It allows the bacteria to uptake environmental DNA across the periplasm and transport it to the cytoplasm, forming a transformant. The environmental DNA initially interacts with a transformation pilus, after which the movement of the DNA into the periplasm is driven by a widely conserved periplasmic protein, ComEA, which binds DNA non-specifically [2, 3]. In Gram-negative bacteria, ComEA contains only a DNA binding domain that freely diffuses in the periplasm [4-6]. It is proposed that ComEA acts as a Brownian ratchet in some Gram-negative bacteria to prevent DNA from diffusing back across the outer membrane [6-8]. In contrast, in Gram-positive bacteria, ComEA is anchored to the membrane by an N-terminal transmembrane domain [2], which is separated from the C-terminal DNA binding domain by ~110 amino acids with an unknown function. This variation in structure suggests it may work differently than the Gram-negative ComEA. To gain insight into the mode of action of Gram-positive *B. subtilis* and *G. stearothermophilus* ComEA ($\mathrm{ComEA_{Bs}}$ and $\mathrm{ComEA_{Gs}}$, respectively), *in vivo* and *in vitro* structure-function studies were performed.

X-ray crystal structures presented later in this chapter and the published manuscript [9] suggested that the ~110 amino acid domain connecting the N- and C- terminal domains was likely responsible for protein oligomerization and therefore termed the oligomerization domain (OD). AUC was then employed to confirm if reversible oligomerization in solution occurred. The effect of the OD domain on DNA binding was also studied using multi-wavelength (MW) AUC, allowing for the determination of the binding ratios. MW-AUC combines the hydrodynamic separation found in sedimentation velocity (SV) experiments with an orthogonal spectral separation component, improving resolution and obtaining information not possible with single wavelength SV experiments. Together with results from x-ray crystal structures, EMSAs, and *in vivo* studies, the mechanism of action of ComEA in Gram-positive bacterial transformations was elucidated.

3.3 Methods

Methods regarding protein preparation, x-ray crystallography, and EMSA can be found in the published paper [9].

3.3.1 Analytical Ultracentrifugation

Sedimentation velocity experiments (SVEs) measure the mass transport of macromolecules in a centrifugal force field in solution and observe the sedimentation and diffusion properties of all species in a mixture, and report their partial concentrations, buoyant molar masses, and shape factors. Sedimentation and diffusion transport in the ultracentrifugation cell are described by the Lamm equation, which is solved using adaptive finite element methods [10, 11]. Whole boundary data obtained in SV experiments are fitted by linear combinations of such solutions using advanced optimization routines [12-14] that are computationally intensive and are carried out on high-performance computing platforms [15]. SVEs were performed in a Beckman Coulter Optima AUC at the Canadian Center for Hydrodynamics at the University of Lethbridge. Data were collected using single- or multiwavelength UV detection. 0.45 ml of sample was filled into double sector epon-charcoal centerpieces equipped with sapphire windows and measured in intensity mode. All experiments were performed at 20 °C, and in a buffer containing 10 mM Tris [pH 8], 100 mM KCl. ComEA$_{Gs}$-OD was measured at 37 krpm. ComEA$_{Gs}$ and DNA were measured at 60 krpm. MW-AUC data involving the 14-bp DNA sequence were measured at 55 krpm, while MW-AUC data involving the 40-bp DNA sequence were collected at 43 krpm. MW-AUC data were recorded in the range of 235–285 nm with 2 nm increments, providing 26 individual datasets for each sample. All data were analyzed using UltraScan 4.0 [16]. The processing of MW-AUC data is described in detail in Henrickson et al., 2022 [17]. Briefly, data from each wavelength were analysed using the two-dimensional spectrum analysis (2DSA) [12], following the workflow described in [18]. After generating time synchronous SVEs, the hydrodynamic profile is spectrally deconvoluted into the molar extinction coefficient profiles of protein and DNA. Buffer density and viscosity corrections were calculated with UltraScan using the partial concentration of each buffer component. Molar extinction profiles were determined by performing separate dilution series for each protein and DNA, collecting an absorbance spectrum across the spectral range of interest (220–300 nm) using a Genesys 10 s benchtop spectrophotometer (Thermo Fisher Scientific). The dilution series of each absorbance spectrum was fitted to an intrinsic extinction profile as described previously [16]. The resulting intrinsic extinction profiles were scaled to molar concentration using an extinction coefficient of 7,450 M^{-1} cm^{-1} at 280 nm for ComEA$_{Gs}$-A108Y and 5,960 M^{-1} cm^{-1} at 280 nm for ComEA$_{Gs}$ (as estimated by UltraScan from protein sequence). Diffusion-corrected sedimentation coefficient profiles were generated using the enhanced van Holde-Weischet analysis implemented in UltraScan [19]. Van Holde-Weischet results are shown as G(s) integral distribution plots, which display the integrated concentration of a sedimenting species on the *y*-axis. As displayed in Figs. 3.2 and 3.3, all concentrations are normalized between 0–100% of the boundary signal. These results suggest a K_d for the monomer-dimer equilibrium between 10 and 160 μM. To determine the dissociation constant for the monomer-dimer equilibrium of ComEA$_{Gs}$, a discrete monomer-dimer model using a genetic algorithm Monte Carlo analysis was fitted to the sedimentation velocity experiment of the intermediate loading concentration (31.3 μM, measured at 280 nm) as described in [20] and [14]. Detailed fitting results are shown in Table 3.2.

3.4 Results

3.4.1 X-ray crystal structures of $\mathrm{ComEA_{Bs}}$ and $\mathrm{ComEA_{Gs}}$

The crystal structures for $\mathrm{ComEA_{Gs}}$ and $\mathrm{ComEA_{Bs}}$ were determined and helped gain insight into the mechanistic function of the ~110 amino acids connecting the N- and C-terminal domains. The crystal structures lead to the conclusion that this previously undescribed domain oligomerized with itself in a head-to-tail orientation, resulting in the authors naming this region the oligomerization domain (OD). Further, it was concluded that the linker region between the OD and DNA-binding domains is flexible, there is minimal contact between the two domains, and the ODs have the potential to form multimers in solution (Figure 3.1). For further findings from X-ray crystallography, please refer to the manuscript [9].

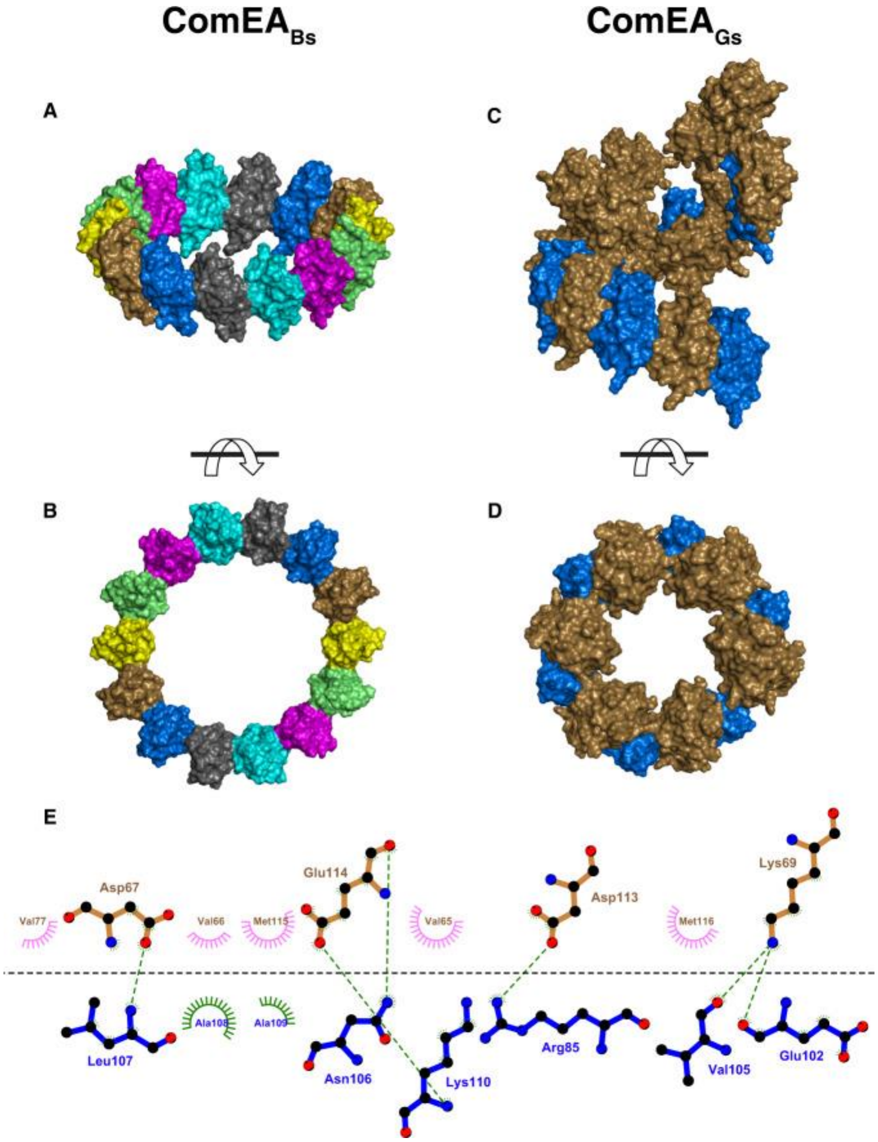


Figure 3.1: ComEA multimerization as observed in the X-ray crystal structures. A and B: Front and side views of the ComEA$_{Bs}$ crystallographic ring, respectively. C and D: Front and side views of the ComEA$_{Gs}$ crystallographic kinked ring, respectively. E: Schematic representation of the ComEA$_{Gs}$ multimerization interface. ComEA chains A and B are depicted as blue and brown bonds, respectively. Hydrogen bonds are depicted as green dashed lines. Hydrophobic contacts are depicted as lines radiating from the semicircles and spheres. The schematic was produced with LigPlot+[21].

3.4.2 ComEA oligomerizes in solution

To confirm that the OD observed in both ComEA X-ray crystal structures mediated reversible self-association in solution, we analyzed different concentrations of ComEA$_{Gs}$ using sedimentation velocity analytical ultracentrifugation (SV-AUC) [20]. At 10.3 μM, the majority of ComEA$_{Gs}$ has a sedimentation coefficient of 1.6 S, consistent with the ComEA monomer. However, at 157 μM, the majority of ComEA$_{Gs}$ has a sedimentation coefficient of 2.2 S, suggesting reversible dimerization (Fig. 3.2A). At an intermediate concentration of 31.3 μM we observed both monomer and dimer, and were able to fit the sedimentation velocity data to a discrete monomer-dimer equilibrium model [20], resulting in a K_d of 33.8 μM (95% confidence intervals: 19.3 μM, 48.4 μM) (Table 3.2). The apparent ComEA dimerization affinity is likely an underestimation of its *in vivo* dimerization affinity, because diffusion of ComEA *in vivo* is limited to two dimensions in the cell membrane.

Consistent with the SV-AUC analysis of ComEA$_{Gs}$, a truncated ComEA$_{Gs}$ protein consisting of the OD alone (ComEA$_{Gs}$-OD) also multimerized in solution (Fig. 3.2B). However, ComEA$_{Gs}$-OD formed primarily monomers and dimers at low concentration (12.7 μM) and larger oligomers with a maximum molar mass around 100 kDa at high concentration (196 μM) (Fig. SF 3.1). One possible kinetic explanation for the different multimeric states of ComEA$_{Gs}$ and ComEA$_{Gs}$-OD at similar concentrations is that the presence of the ComEA$_{Gs}$ DNA-binding domain slows the OD multimerization search, limiting collisions, thus slowing the OD multimerization rate (k_{on}). To confirm that ComEA multimerization is driven by the OD, we examined the X-ray crystal structures to identify a small amino acid buried in the OD multimerization interface that we could replace with a larger amino acid, introducing steric bulk and disrupting multimerization. We identified ComEA$_{Gs}$ Ala108 (corresponding to ComEA$_{Bs}$ Ala106) and substituted it with tyrosine (Figs. 3.1E, 3.2A, C, and Table 3.2). Indeed, ComEA$_{Gs}$-A108Y was monomeric at both 11.3 μM and 124 μM (Fig. 3.2A). These SV-AUC experiments demonstrate that the OD drives ComEA multimerization in solution as predicted from the X-ray crystal structures.

3.4.3 ComEA oligomerizes on DNA

We further speculated that ComEA multimerization plays an important role in its interaction with DNA. To test this, we used multi-wavelength analytical ultracentrifugation (MW- AUC) [17, 22-24] to measure the binding of wild-type ComEA$_{Gs}$ and ComEA$_{Gs}$-A108Y to DNA. MW-AUC is a novel technique that permits the spectral separation of protein and DNA species based on their unique absorbance spectra. Consequently, it is possible to measure the molar stoichiometry of the complexes formed and identify the type of macromolecule(s) forming each hydrodynamic species. The hydrodynamic measurements depend on each species' molar mass and hydrodynamic radius.

To elucidate the interactions between ComEA and double-stranded DNA, we mixed wild-type and mutant ComEA with a 14-bp double-stranded DNA duplex at a 5:1 and a 10:1 molar protein excess and a 10:1 protein excess over a 40-bp double-stranded DNA duplex (Fig. 3.3A–C and Table 3.1). In all cases, the DNA concentration was held constant at 1.5 μM, and a well- defined complex was formed, with less than 10% of the DNA remaining free in the solution. The deconvoluted DNA sedimentation pattern suggests the presence of a saturated, homogeneous complex being formed in all cases. As expected, the DNA sequestered excess protein until it was fully occupied, and the 14-bp sequence accommodated fewer ComEA molecules than the 40-bp sequence.

The sedimentation coefficient of the complex differed for wildtype and mutant and depended on the length of the DNA. The 14-bp double-stranded DNA duplex sedimented at 2.0 S in isolation but formed a 3.3 S complex with ComEA$_{Gs}$-A108Y, and a 3.9 S complex with wild-type ComEA$_{Gs}$ in the 5:1 mixture (Fig. 3.3A). For this mixture, we measured a ~3:1 protein:DNA ratio for the complex formed with ComEA$_{Gs}$, and a ~2:1 ratio for the complex formed with the ComEA$_{Gs}$A108Y protein. For the 10:1 protein:DNA mixture, a slightly larger 3.5 S complex was formed with ComEA$_{Gs}$-A108Y, a 4.0 S complex was formed with wild-type ComEA$_{Gs}$, and the protein:DNA ratios did not change significantly (Fig. 3.3B).

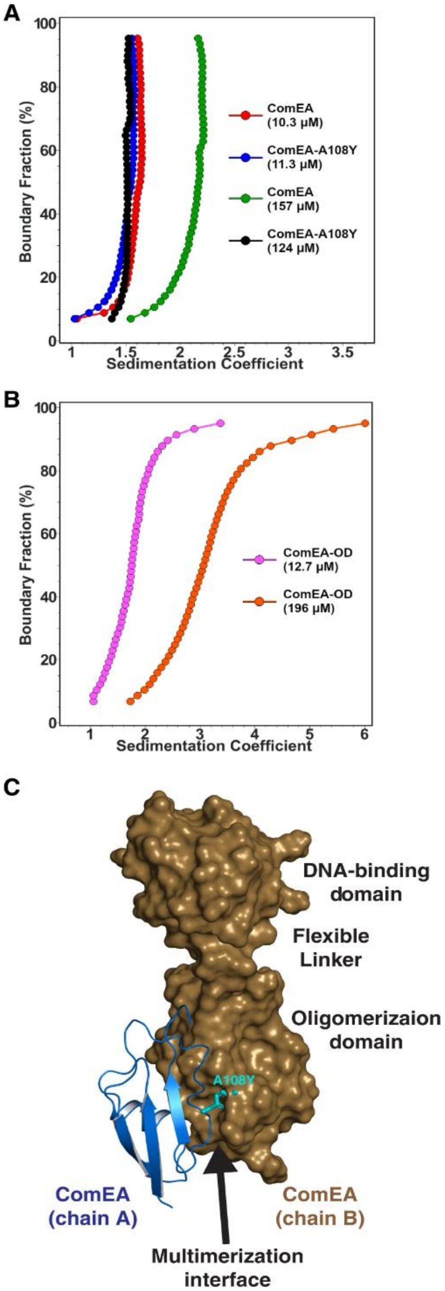


Figure 3.2: AUC analysis of ComEA, ComEA-A108Y, and the ComEA oligomerization domain. A: Integral sedimentation coefficient distribution overlays comparing the dimerization potential of ComEA$_{Gs}$ at 10.3 µM (red) and 157 µM (green) and ComEA$_{Gs}$-A108Y at 11.3 µM (blue) and 124 µM (black). Only ComEA$_{Gs}$ dimerizes at higher concentration, while ComEA$_{Gs}$-A108Y remains monomeric. B: Integral sedimentation coefficient distribution overlays of the ComEA$_{Gs}$ OD at 12.7 µM (magenta) and 196 µM (orange), showing reversible self-association. C: Structure of ComEA$_{Gs}$ with chain A depicted as a cartoon and chain B depicted as a surface. ComEA$_{Gs}$ chain A Ala108 was mutated to Tyr and is depicted as cyan sticks

The 40-bp DNA sediments at 3.3 S when examined by itself, but, when mixed in a 10:1 ratio with ComEA$_{Gs}$-A108Y resulted in a 6.3 S complex, while the same mixture with wild-type ComEA$_{Gs}$ resulted in an 8.1 S complex (Fig. 3.3C). For the 40-bp DNA, the difference between protein:DNA molar ratios of the formed complexes was more pronounced between wild-type and mutant. The wild-type ComEA$_{Gs}$ displayed an approximately 7:1 ratio, but the complex between DNA and ComEA$_{Gs}$-A108Y suggested the presence of about a 4.7:1 ratio. While the accuracy of these ratios depends on the estimated molar extinction coefficients in the formed complexes, there is a clear difference in the amount of ComEA bound to DNA for the wild-type and mutant proteins.

In sum, these results suggest that wild-type ComEA, capable of oligomerizing, forms larger complexes with DNA than the A108Y mutant. Oligomerization facilitates efficient and cooperative packing of ComEA on DNA, which is further supported by gel shift analysis (detailed in the manuscript [9]).

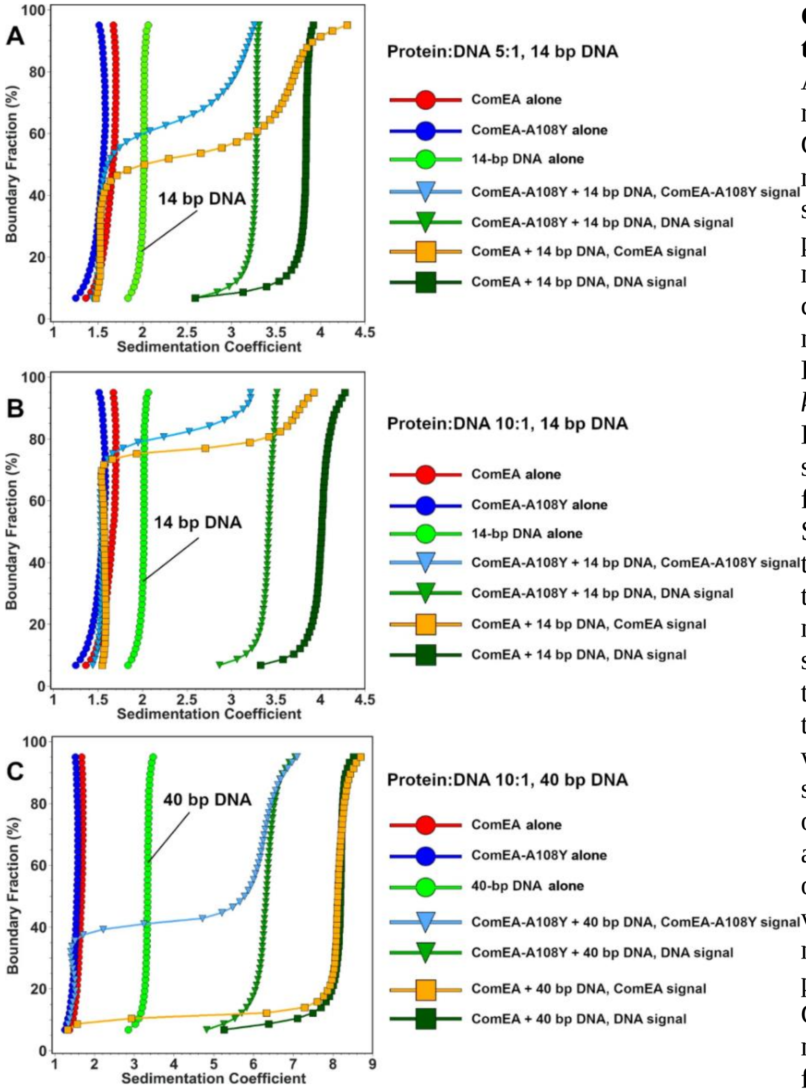


Figure 3.3: Fig. 5 | AUC analysis of ComEA and ComEA-A108Y, and their interactions with 1.5 µM DNA. A and B: Mixtures 5:1 and 10:1 molar ratios of wild-type ComEA$_{Gs}$ and ComEA$_{Gs}$-A108Y with the 14-bp DNA molecule, respectively. Here, the DNA signals still suggest full saturation with protein, but the protein signals show more heterogeneous sedimentation coefficient distributions, consistent with more rapid exchange with the protein-DNA complex, which suggests a faster k_{off} rate constant. More than 90% of the DNA is complexed with protein, shifting the 14 bp DNA distribution from 2.0 S for the control by itself to 3.3 S for the mutant, and 3.9 S for the wild-type in panel A. In panel B, increasing the protein concentration to 10:1, marginally shifts the DNA sedimentation for the wild-type further to about 4.0 S, while barely affecting the DNA sedimentation when mixed with the mutant. C: Integral sedimentation coefficient distribution overlays for the deconvoluted protein and DNA signals from the 10:1 mixture of ComEA$_{Gs}$ and ComEA$_{Gs}$-A108Y with 1.5 µM of the 40-bp DNA molecule. Unbound ComEA in the presence of DNA co-sediments with ComEA in the absence of DNA. Again, more than 90% of the DNA signal shifts from the position of free 40 bp DNA at 3.3 s to a homogeneous composition at 6.3 S for the mutant, and 8.1 S for the wild-type, suggesting saturation of the DNA with ComEA. The ComEA signal closely tracks the DNA signal, suggesting a tight complex formation with a slow k_{off} rate constant. For all plots, reference controls of each molecule by itself are shown as circles (ComEA: red circles, ComEA-A108Y: blue circles, 14 or 40 bp DNA, as indicated: green circles), symbols for interactions between DNA and wildtype protein are shown as squares (ComEA signal: orange squares, DNA signal: dark green squares) and interactions between ComEA-A108Y and DNA are shown in triangles (ComEA-A108Y signal: olive triangles, DNA signal: light blue triangles).

3.5 Discussion

This study presented several notable findings. The first was the identification of the oligomerization domain (OD) of ComEA in *B. subtilis* and *G. stearothermophilus* by X-ray crystallography. AUC was then used to confirm that the OD was the driving force of oligomerization and to determine the k_d of oligomerization. When the OD was isolated and investigated via AUC, it was also found that it could form extended multimers. The second finding used MW-AUC to determine wild-type and ComEA-A108Y molar binding ratios to DNA. The molar binding ratios to the 40 bp DNA were 7:1 for the wild-type ComEA and 4.7:1 for ComEA-A108Y. This difference in molar ratio likely arises from the OD, which can bring additional ComEAs into close proximity to the DNA, leading to a cooperative binding effect, as further supported by the EMSA results shown in the manuscript. Finally, the *in vivo* studies demonstrated the importance of the OD domain for the uptake of DNA and that mutations or removal of the OD domain result in significant reductions in uptake, likely signifying that in Gram-positive bacteria, ComEA does more than act as a Brownian ratchet mechanism and may be important for the condensation of the DNA within the periplasm.

This study highlighted the ability of AUC to study pure samples; it helped elucidate the function of ComEA's OD and was used to determine the oligomerization K_d and k_{off} rates, to be 3.38 e^{-5} M and 5.27 e^{-5} M (Table 3.2). MW-AUC was used to study the function of the OD domain in relation to ComEA-DNA binding. The binding ratio of ComEA to DNA was determined through optical deconvolutions, resulting in the unambiguous identification of DNA and protein, highlighting the binding differences between ComEA and ComEA-A108Y. This study demonstrates the power of MW-AUC to enhance the characterization of interacting molecules.

3.6 References

1. Dubnau D, Blokesch M. Mechanisms of DNA Uptake by Naturally Competent Bacteria. Annu Rev Genet. 2019;53:217-237. Epub 2019/08/23. doi: 10.1146/annurev-genet-112618-043641. PubMed PMID: 31433955.

2. Inamine GS, Dubnau D. ComEA, a Bacillus subtilis integral membrane protein required for genetic transformation, is needed for both DNA binding and transport. J Bacteriol. 1995;177(11):3045-3051. Epub 1995/06/01. doi: 10.1128/jb.177.11.3045-3051.1995. PubMed PMID: 7768800; PubMed Central PMCID: PMCPMC176991.

3. Provvedi R, Dubnau D. ComEA is a DNA receptor for transformation of competent Bacillus subtilis. Mol Microbiol. 1999;31(1):271-280. Epub 1999/02/13. doi: 10.1046/j.1365-2958.1999.01170.x. PubMed PMID: 9987128.

4. Gangel H, Hepp C, Müller S, Oldewurtel ER, Aas FE, Koomey M, et al. Concerted spatio-temporal dynamics of imported DNA and ComE DNA uptake protein during gonococcal transformation. PLoS Pathog. 2014;10(4):e1004043. Epub 2014/04/26. doi: 10.1371/journal.ppat.1004043. PubMed PMID: 24763594; PubMed Central PMCID: PMCPMC3999279.

5. Seitz P, Blokesch M. DNA transport across the outer and inner membranes of naturally transformable Vibrio cholerae is spatially but not temporally coupled. mBio. 2014;5(4). Epub 2014/08/21. doi: 10.1128/mBio.01409-14. PubMed PMID: 25139903; PubMed Central PMCID: PMCPMC4147865.

6. Seitz P, Pezeshgi Modarres H, Borgeaud S, Bulushev RD, Steinbock LJ, Radenovic A, et al. ComEA is essential for the transfer of external DNA into the periplasm in naturally transformable Vibrio cholerae cells. PLoS Genet. 2014;10(1):e1004066. Epub 2014/01/07. doi: 10.1371/journal.pgen.1004066. PubMed PMID: 24391524; PubMed Central PMCID: PMCPMC3879209.

7. Hepp C, Maier B. Kinetics of DNA uptake during transformation provide evidence for a translocation ratchet mechanism. Proc Natl Acad Sci U S A. 2016;113(44):12467-12472. Epub 2016/11/03. doi: 10.1073/pnas.1608110113. PubMed PMID: 27791096; PubMed Central PMCID: PMCPMC5098643.

8. Peskin CS, Odell GM, Oster GF. Cellular motions and thermal fluctuations: the Brownian ratchet. Biophys J. 1993;65(1):316-324. Epub 1993/07/01. doi: 10.1016/s0006-3495(93)81035-x. PubMed PMID: 8369439; PubMed Central PMCID: PMCPMC1225726.

9. Ahmed I, Hahn J, Henrickson A, Khaja FT, Demeler B, Dubnau D, et al. Structure-function studies reveal ComEA contains an oligomerization domain essential for transformation in gram-positive bacteria. Nat Commun. 2022;13(1):7724. Epub 2022/12/14. doi: 10.1038/s41467-022-35129-0. PubMed PMID: 36513643; PubMed Central PMCID: PMCPMC9747964.

10. Cao W, Demeler B. Modeling analytical ultracentrifugation experiments with an adaptive space-time finite element solution of the Lamm equation. Biophys J. 2005;89(3):1589-1602. Epub 2005/06/28. doi: 10.1529/biophysj.105.061135. PubMed PMID: 15980162; PubMed Central PMCID: PMCPMC1366663.

11. Cao W, Demeler B. Modeling analytical ultracentrifugation experiments with an adaptive space-time finite element solution for multicomponent reacting systems. Biophys J. 2008;95(1):54-65. Epub 2008/04/09. doi: 10.1529/biophysj.107.123950. PubMed PMID: 18390609; PubMed Central PMCID: PMCPMC2426643.

12. Brookes E, Cao W, Demeler B. A two-dimensional spectrum analysis for sedimentation velocity experiments of mixtures with heterogeneity in molecular weight and shape. Eur Biophys J. 2010;39(3):404-414.

13. Brookes E, Demeler B. Parsimonious Regularization Using Genetic Algorithms Applied to the Analysis of Analytical Ultracentrifugation Experiments. GECCO ’07: Proceedings of the 9th Annual Conference on Genetic and Evolutionary Computation, vols. 361–368, July 2007. ACM 978-1-59593-697-4/07/0007. 2007. doi: https://doi.org/10.1145/1276958.1277035.

14. Demeler B, Brookes E. Monte Carlo analysis of sedimentation experiments. Colloid and Polymer Science. 2008;286:129-137.

15. Brookes E, Demeler B. Parallel computational techniques for the analysis of sedimentation velocity experiments in UltraScan. Colloid and Polymer Science. 2008;286:139-148.

16. Demeler B, Gorbet GE. In Analytical Ultracentrifugation: Instrumentation, Software, and Applications (eds Uchiyama S., Arisaka F., Stafford W. F. & Laue T.) 119-143 (Springer Japan). 2016. doi: DOI: 10.1007/978-4-431-55985-6_8.

17. Henrickson A, Gorbet GE, Savelyev A, Kim M, Hargreaves J, Schultz SK, et al. Multi-wavelength analytical ultracentrifugation of biopolymer mixtures and interactions. Anal Biochem. 2022:114728.

18. Demeler B. Methods for the design and analysis of sedimentation velocity and sedimentation equilibrium experiments with proteins. Current protocols in protein science. 2010;60(1):7.13. 11-17.13. 24.

19. Demeler B, Van Holde KE. Sedimentation velocity analysis of highly heterogeneous systems. Anal Biochem. 2004;335(2):279-288.

20. Demeler B, Brookes E, Wang R, Schirf V, Kim CA. Characterization of reversible associations by sedimentation velocity with UltraScan. Macromol Biosci. 2010;10(7):775-782. Epub 2010/05/21. doi: 10.1002/mabi.200900481. PubMed PMID: 20486142.

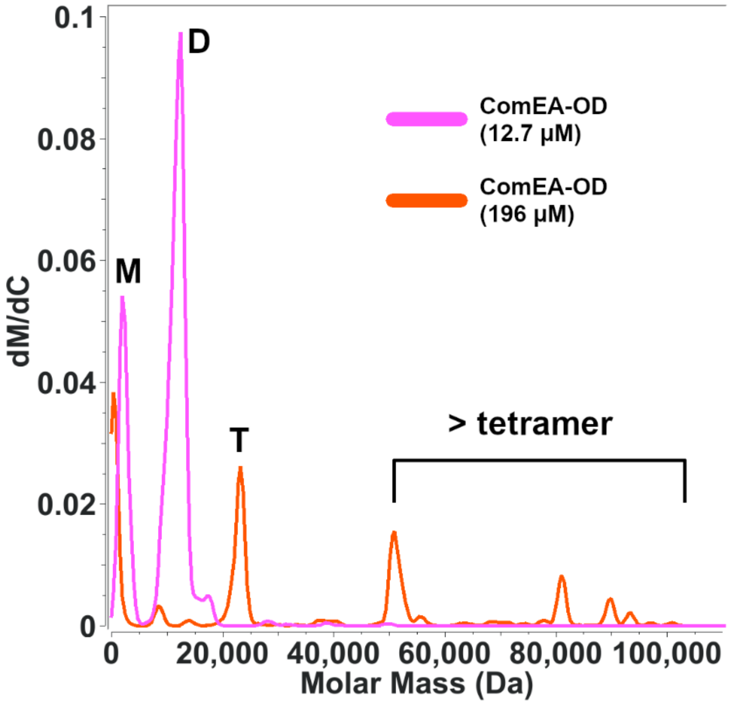
21. Laskowski RA, Swindells MB. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. J Chem Inf Model. 2011;51(10):2778-2786. Epub 2011/09/17. doi: 10.1021/ci200227u. PubMed PMID: 21919503.

22. Horne CR, Henrickson A, Demeler B, Dobson RC. Multi-wavelength analytical ultracentrifugation as a tool to characterise protein–DNA interactions in solution. Eur Biophys J. 2020;49(8):819-827.

23. Demeler B. Measuring molecular interactions in solution using multi-wavelength analytical ultracentrifugation: combining spectral analysis with hydrodynamics. Biochem (Lond). 2019;41(2):14-18. Epub 2019/01/01. doi: 10.1042/bio04102014. PubMed PMID: 32952314; PubMed Central PMCID: PMCPMC7500494.

24. Gorbet GE, Pearson JZ, Demeler AK, Cölfen H, Demeler B. Next-generation AUC: analysis of multiwavelength analytical ultracentrifugation data. Methods Enzymol. 562: Elsevier; 2015. p. 27-47.

3.7 Supplemental Information



SF 3.1: Molar mass distribution of ComEA$_{Gs}$-OD at the same concentrations as shown in Fig. 3.2B.

Table 3.1. Oligonucleotides used in this study

Name	Sequence
His6-BsuF	5’- TAAGAAGGAGATATACCATGTCAAACGAAACAATTGTGATAGACATC -3’
His6-BsuF	5’- GTGGTGGTGGTGGTGGTGCTCGAGCTTTACTGTAATGGAAGACTTTATTTTCTCAA -3’
His-Sumo-Gsth_F	5’- AGAGAACAGATTGGTGGTGCGTCAAAAACCGCTGTGGTGG -3’
His-Sumo-Gsth_R	5’- GTCACCCGGGCTCGAGTTACGGAACAAGCAAATACGGC -3’
His-Sumo-OD_R	5’- GTCACCCGGGCTCGAGTTATTTTTTCGGCACGTA -3’
EAGsA108Y Fwd	5’- CGAAAGTCAATTTGTACGCAAAAGTGCATGATGA -3’
EAGsA108Y Rev	5’- TTTCATCCGCCTCATCCGTCAATC -3’
EAGsK166A Fwd	5’- GGCATCGGACCGGCGGCGGCAAACGCGATCATCGCC -3’
EAGsK166A Rev	5’- GGGAAGCTGCATCAACTCCTCTTCCGTCGCCGTATTGAT -3’
EAGsK201A Fwd	5’- CTGGAGAAATTGGCACCGTATTTGCTTGTT -3’
EAGsK201A Rev	5’- CGTTTTTTCACCAATTCCAGTCACATTCAACA -3’
EABsK164A Fwd	5’- GGGGCCATCCGCTGCTGAAGCTATTATTG -3’
EABsK164A Rev	5’- ACCCCTGAGATGCCTTGT -3’
EABsK193A Fwd	5’- AATAGGTGAAGCTTCATTTGAGAAAATAAAGTCTTCC -3’
EABsK193A Rev	5’- CCTGAAACCTTAGTGATATC -3’
EABsK197A Fwd	5’- GTCATTTGAGGCTATAAAGTCTTCCATTAC-3’
EABsK197A Rev	5’- TTTTCACCTATTCCTGAAAC -3’
EABsK199A Fwd	5’- TGAGAAAATAGCTTCTTCCATTACAGTAAAGTGAATG -3’
EABsK199A Rev	5’- AATGACTTTTCACCTATTCC -3’
EABsA106Y Fwd	5’- AGTAAATTTGTATGAGATTCTGCAGGACGGGAC -3’
EABsA106Y Rev	5’- TGCGCTTCGTCTGCTTGT -3’
EABsD111N Fwd	5’- GATTCTGCAGAATGGGACAGTGG -3’
EABsD111N Rev	5’- TCCGCCAAATTTACTTGC -3’
EABsΔOD Fwd	5’- GGAGAGGAAACAGCAGTG -3’
EABsΔOD Rev	5’- TGACTCGTCGTTGTTTGC -3’
EAGs30bp Top	5’- GCACTGTGAATATCACTACAGCAACCTTAG -3’
EAGs30bp Bot	5’- CTAAGGTTGCTGTAGTGATATTCACAGTGC -3’
EAGs40bp Top	CGTTAGCACTGTGAATATCACTACAGCAACCTTAGAATCG
EAGs40bp Bot	CGATTCTAAGGTTGCTGTAGTGATATTCACAGTGCTAACG
EAGs14bp Top	GTAGGACCATCTGC
EAGs14bp Bot	GCAGATGGTCCTAC

Table 3.2: Discrete Model Genetic Algorithm (MC) fit for 31.3 µM ComEA$_{GS}$ sedimentation velocity experiment.

Model:	Monomer (1) – Dimer (2) reversibly self-associating
Residual RMS Deviation:	0.00134989 absorbance units (280 nm)
Weight Average S20,W:	1.6291e-13 s
Weight Average D20,W:	1.2290e-06 cm2/s
Total Concentration:	2.2372e-01 absorbance units (280 nm)
Fitted partial specific volume:	0.795 (ml/g)

Distribution Information:

	Molec. Wt. (kDalton)	S20,W (s)	D20,W (cm2/s)	f/f0
Monomer:	15.6	1.6291e-13	1.2290e-06	1.03
Dimer:	31.2	1.9611e-13	7.3972e-07	1.35

Discrete Model GA-MC Summary Statistics (96 Monte Carlo iterations):

Component	Attribute	Mean_Value	95%_Confidence(low)	95%_Confidence(high)
(All)	RMSD	1.3482e-03	1.3376e-03	1.3589e-03
Monomer	Concentration (AU, 280 nm)	2.2372e-01	2.2348e-01	2.2395e-01
Monomer	Partial specific volume (ml/g)	7.9491e-01	7.8601e-01	8.0381e-01
Monomer	Molecular weight (Da)	1.5646e+04	(Fixed)	
Monomer	Sedimentation coefficient (s)	1.6292e-13	1.5939e-13	1.6645e-13
Monomer	Diffusion coefficient (cm2/s)	1.2295e-06	1.1841e-06	1.2750e-06
Monomer	Frictional ratio	1.0245e+00	9.8290e-01	1.0661e+00
Monomer	Concentration (AU, 280 nm)	2.2372e-01	2.2348e-01	2.2395e-01
2	Partial specific volume (ml/g)	7.9491e-01	7.8601e-01	8.0381e-01
2	Molecular weight (Da)	3.1292e+04	(Fixed)	
2	Sedimentation coefficient (s)	1.9610e-13	1.9303e-13	1.9917e-13
2	Diffusion coefficient (cm2/s)	7.3994e-07	7.1686e-07	7.6302e-07
2	Frictional ratio	1.3510e+00	1.3042e+00	1.3978e+00

Reversible Associations Information:

Attribute	Mean_Value	95%_Confidence (low)	95%_Confidence (high)
K_d (M)	3.3843e-05	1.9287e-05	4.8400e-05
k_{off} Rate (1/s)	5.2672e-05	1.6950e-05	8.8395e-05

Chapter 4. Density matching multi-wavelength analytical ultracentrifugation to measure drug loading of lipid nanoparticle formulations

Amy Henrickson[1], Jayesh A. Kulkarni[2], Josh Zaifman[2], Gary E. Gorbet[3], Pieter R. Cullis[2], and Borries Demeler[1,3,4]*

[1] Department of Chemistry and Biochemistry, The University of Lethbridge, Lethbridge, Alberta, Canada T1K 3M4

[2] Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z2

[3] AUC Solutions, Houston, Texas 77494, United States

[4] Department of Chemistry, University of Montana, Missoula, Montana 59812, United States

4.1 Contributions of Authors

This chapter is adapted from a manuscript published in *ACS Nano,* Volume 15(3), pp. 5068-5076, on February 22, 2021 (https://doi.org/10.1021/acsnano.0c10069). Dr. Jayesh Kulkarni and I share first author-ship. I performed and analyzed all AUC experiments and assisted with the experimental design and writing of the manuscript with Dr. Borries Demeler, who also conceived the project. Dr. Jayesh Kulkarni, Dr. Josh Zaifman, and Dr. Pieter Cullis prepared and contributed the LNPs and performed TEM and DLS measurements, and Dr. Gary Gorbet developed the density matching software in UltraScan.

4.2 Abstract

Previous work suggested that lipid nanoparticle (LNP) formulations, encapsulating nucleic acids, display electron-dense morphology when examined by cryogenic transmission electron microscopy (cryo-TEM). Critically, the employed cryo-TEM method cannot differentiate between loaded and empty LNP formulations. Clinically relevant formulations contain high lipid-to-nucleic acid ratios (10−25 (w/w)), and for systems that contain mRNA or DNA, it is anticipated that a substantial fraction of the LNP population does not contain a payload. Here, we present a method based on the global analysis of multi-wavelength sedimentation velocity analytical ultracentrifugation, using density matching with heavy water, that not only measures the standard sedimentation and diffusion coefficient distributions of LNP mixtures, but also reports the corresponding partial specific volume distributions and optically separates signal contributions from nucleic acid cargo and lipid shell. This makes it possible to reliably predict molar mass and anisotropy distributions, in particular, for systems that are heterogeneous in partial specific volume and have low to intermediate densities. Our method makes it possible to unambiguously measure the density of nanoparticles and is motivated by the need to characterize the extent to which lipid nanoparticles are loaded with nucleic acid cargoes. Since the densities of nucleic acids and lipids substantially differ, the measured density is directly proportional to the loading of nanoparticles. Hence, different loading levels will produce particles with variable density and partial specific volume. An UltraScan software module was developed to implement this approach for routine analysis.

4.3 Keywords

lipid nanoparticles, multi-wavelength analytical ultracentrifugation, density matching, RNA gene therapy, nanomedicine, lipid biophysics, loading heterogeneity

4.4 Introduction

Analytical ultracentrifugation (AUC) is a solution-based separation technique with nearly a century of applications in biomolecular research. It is a first principles technique which does not require external standards and has recently seen significant advances in instrumentation and analysis software [1-11]. During sedimentation velocity experiments (SVEs), centrifugal forces create moving concentration gradients, and their evolution over time and space are monitored by either their absorbance at one or more wavelengths [6], their fluorescence emission (if the molecules contain fluorophores), or their refractive index changes, using Rayleigh interference. Mass transport of the solutes occurs through both sedimentation and diffusion. Both transport processes are impacted by the frictional properties of the molecules and the viscosity of the solvent through molecular surface−solvent interactions. Sedimentation transport also depends on the mass of the sedimenting particles, their partial specific volumes (PSVs), and the density of the solvent. All experimental data can be directly fitted to finite element solutions of partial differential equations [12, 13], providing hydrodynamic details such as sedimentation and diffusion coefficients, as well as partial concentrations of analytes present in mixtures. Sedimentation and diffusion coefficients can further be interpreted to derive anisotropy, density, and molar mass, as well as thermodynamic interaction parameters when studying mass action effects. In this study, we employ SVEs performed in multi-wavelength mode (MW-AUC) [14-16], coupled to density matching with heavy water, to characterize the loading of lipid nanoparticles (LNPs) [17].

LNPs represent a mature drug-delivery technology that enable the therapeutic potential of nucleic acids. Onpattro, an RNA interference therapeutic approved by the U.S. Food and Drug Administration, uses LNPs to entrap short interfering RNA (siRNA), protects it from degradation during circulation, and enhances intracellular delivery into hepatocytes at the liver. These LNP-siRNA are approximately 50 nm in diameter [18] and are electron-dense when visualized by cryogenic-transmission electron microscopy (cryo-TEM) [19]. A critical knowledge gap is the inability to differentiate between empty LNPs and those loaded with nucleic acid by using the same cryo-TEM technique. One exception is the use of formulations that contain completely charge-neutralized ionizable lipids where clear multilamellar structures are observed. Comparatively, clinically relevant LNP contain substantially higher amounts of ionizable lipid (amine-tophosphate (N/P) ratios of 3−6), resulting in indistinguishable morphology between empty and partially loaded LNPs. Likewise, the interparticle distribution of siRNA cannot be easily determined. The presence of siRNA can be detected using spectrophotometric methods, but they do not differentiate loaded from free nucleic acid. Dynamic light scattering (DLS) provides information about the size of particles, but similarly sized LNPs containing nucleic acids cannot be resolved from those without cargo. Hence, such methods are unable to distinguish heterogeneity of loading and fail to distinguish empty from partially filled LNPs.

Determining the exact amount of nucleic acid that is loaded into an LNP poses a complicated challenge for several reasons: The self-assembly of these particles invariably leads to some heterogeneity in the particle size (even with low polydispersity indices of <0.1) and interparticle RNA copy number. Their physicochemical analyses are typically performed on bulk samples and represent an average of the population. Solution techniques such as dynamic light scattering and size exclusion chromatography have difficulties distinguishing between LNPs which are empty or partially or fully loaded with cargo due to limitations in resolution and separability. It is important to note that the two components in LNPs (lipid and nucleic acid) have very different densities. Nucleic acids are much denser than the lipid excipients, which include four components: an ionizable cationic lipid, phospholipid, cholesterol, and poly-(ethylene glycol) (PEG) lipid. The method presented here exploits this difference in density to estimate the loading amounts of LNPs by employing a MW-AUC approach that is based on density matching with heavy water. At the same time, this method leverages the spectral differences between a lipid nanostructure, which scatters light, and the nucleic acid cargo, which has a unique chromophore at ∼260 nm, to achieve optical separation, orthogonal to the hydrodynamic separation. As we demonstrate here, this method is ideally suited for the solution composition characterization of LNP preparations, which are inherently heterogeneous in size and loading. Unlike electron microscopy and single-molecule techniques, it provides high statistical certainty through bulk observation. A multi-wavelength detector offers differential detection of samples containing a mixture of analytes that vary in chemical composition. Our method is sensitive to all parameters of interest for LNP characterization and, by virtue of the extensions discussed here, successfully addresses the challenges posed by LNP formulation characterization.

4.5 Results and Discussion

We investigated the loading of LNPs with siRNA by multiwavelength AUC (MW-AUC) and density matching SVEs. Combining these two methods by global analysis produces orthogonal information, which enhances the confidence in the results obtained by either method alone and provides very high resolution and detail on particle size and cargo loading. To illustrate this approach, we compared LNP preparations containing siRNA to empty LNPs, free nucleic acid, and protein. To further validate the method, hydrodynamic radius (Rh) distributions obtained from MW-AUC were compared to those obtained by cryo-TEM and dynamic light scattering. Three key preparations are referenced in the following text: empty LNPs, and LNP-siRNA systems generated at a loading ratio of 1 or 6 molecules of ionizable cationic lipid for each nucleic acid phosphate, referred to as NP1 or NP6, respectively.

4.5.1 Analysis of Partial Specific Volume.

Density variations in the LNP preparations are assessed by measuring their partial specific volume (PSV) using a density matching approach [17]. The PSV can be viewed as the inverse of the particle density of the hydrated, sedimenting particle. The PSV results for these LNP preparations, together with reference standards, including siRNA used for encapsulation, a 2.88kb double-stranded DNA fragment, and bovine serum albumin protein (BSA), are shown in Figure 4.1. In Figures 4.1−4.5, and Figure 4.9, we report distributions using integral representations, which directly provide fractions of the total concentration on the y axis for different species. For example, in Figure 4.1, all of the plasmid DNA (black line) and siRNA exhibit the same PSV for all fractions, producing a vertical plot, which suggests homogeneity in PSV. BSA (magenta line) shows two discrete PSV values for monomer (0.68 mL/g, ∼80%) and dimer (0.66 mL/ g, ∼20%), and LNPs (red, green, and blue lines) produce heterogeneous distributions, where each fraction relates to a different PSV value.

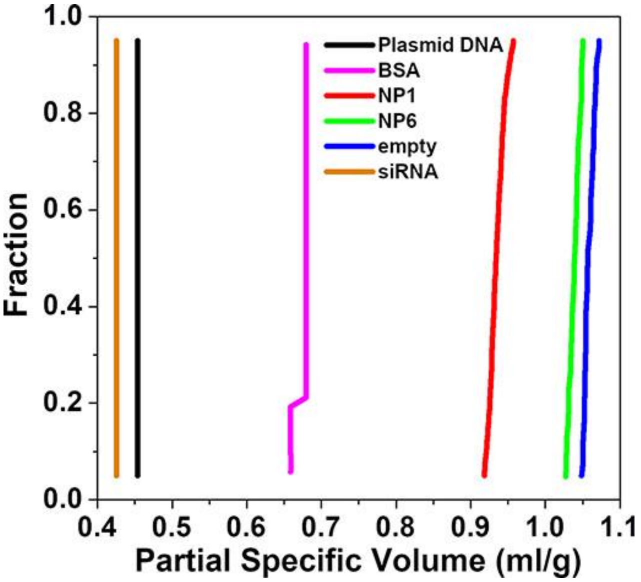


Figure 4.1: PSV distributions for pure 21 bp siRNA (brown), 2.88 kb linear double-stranded plasmid DNA (black), bovine serum albumin (BSA, magenta), NP1 (red), NP6 (green), and empty LNP (blue).

As expected, empty LNPs display the lowest density distribution (highest partial specific volume), with average PSV values of 1.06 mL/g, consistent with their flotation in light water. Empty LNPs are followed by NP6 LNPs, resulting in an average PSV of 1.04 mL/g, still slightly higher than light water. LNPs with loading ratio NP1 display an average PSV of 0.935 mL/g, which is less than the PSV of light water (1.0 mL/g), causing sedimentation in light water. The PSV values for BSA (monomer at 0.68 mL/g, ~80%, and dimer at 0.66 mL/g, ~20%) are slightly lower than predicted from sequence (0.733 mL/g, as implemented in UltraScan [20]), which could be explained by hydrogen-deuterium exchange (HDX) that will occur on exposed side chains during the incubation in D_2O required for the density matching experiments. The lowest PSVs of 0.43 and 0.46 mL/g are observed for methylated and double-stranded siRNA and plasmid DNA, respectively, when sedimenting in a buffer containing 150 mM NaCl. The PSV values observed for pure nucleic acids are also lower than expected due to HDX [21]. Using the known molar mass for the siRNA (13529 Da) and the measured sedimentation and diffusion coefficients in light water buffer ($s_{20,\mathrm{W}}$, 2.58 s; $D_{20,\mathrm{W}}$, 9.54 cm^2/s), the calculated PSV is 0.514

mL/g for siRNA. Whenever absolute PSV values must be measured, which is not required for the purposes of this study, we recommend to substitute D_2O by $H_2{}^{18}O$. Of note, the PSV distributions of all three LNP preparations indicate a slight heterogeneity in the PSV, suggesting heterogeneity in particle size and composition (zoomed-in portion in Supplemental Information Figure SF 4.1 for clarity). Figure SF 4.1 also shows that there is minimal, if any, overlap in PSV between NP6 and empty LNPs, and none with NP1, indicating that neither NP1 nor NP6 contain any measurable empty LNP components. Importantly, neither loading ratio suggests the presence of free nucleic acid, whose PSV would be <0.5 mL/g according to the siRNA and DNA density matching controls. Clear differences observed in the PSV distributions of different loadings of LNPs therefore provide a sensitive approach to compare LNP loading states. Though it is not immediately obvious why empty LNPs display a slight heterogeneity in PSV, a possible explanation is that PEG molecules, which are positioned on the surface of the LNP, are more hydrated. Since water has a higher density than the lipids on the interior of the LNP, a variation in size will therefore affect the ratio of less dense lipids in the core *vs* hydrated and therefore higher density moieties in the shell. This will affect the overall density of the particles, and therefore a mixture of differently sized empty LNPs will display a range of PSV values. Our results show clear differences in PSV for the three tested LNPs, reflecting the composition and level of nucleic acid load of these samples.

4.5.2 Analysis of Molar Mass and Particle Size Distributions

Once a PSV value has been determined, it is possible to derive distributions for molar mass and R_h from AUC results. Molar mass and R_h for the three LNPs are shown in Figure 4.2. As expected, empty LNPs exhibit the lowest molar mass and the smallest R_h of the three preparations, while NP6 and NP1 demonstrate increasing mass and R_h, respectively.

Heterogeneity in the distributions is again apparent in both mass and R_h, which correlates with the heterogeneity observed in the partial specific volume (see Figure SF 4.1).

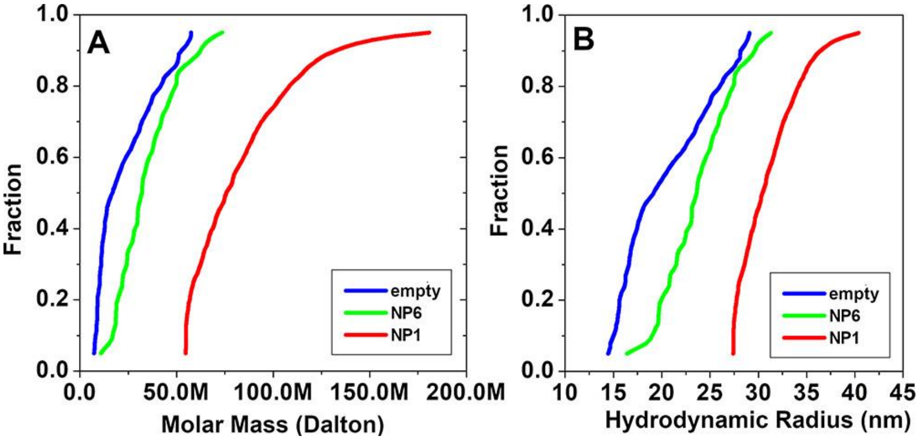


Figure 4.2: Molar mass distributions (A) and hydrodynamic radii distribution (B) of empty (blue), NP6 (green), and NP1 (red) LNP preparations.

Particle sizing by cryo-TEM corroborates R_h distributions observed by AUC (see Figure 4.3). In each case, the cryo-TEM data suggest slightly smaller radii than those observed by AUC, which can be explained by the fact that the radii measured by cryo-TEM are based on the electron-dense regions of the particle, not on the hydrodynamic radius, which is measured by AUC or DLS (see Figure 4.4). The hydrodynamic radius includes the hydration layer, as well as the PEG chains, which extend from the lipid core into the solvent, but have insufficient electron density to be visualized by cryo-TEM; a white ring that appears around the outside of LNPs that has been attributed to PEG lipids is actually the result of a defocused image [22, 23]. The shapes of the R_h distributions of AUC and cryo-TEM differ slightly, which can be attributed to the fact that AUC does not count individual particles but is based on bulk observation and, hence, has more reliability due to better counting statistics. Whether measured by AUC or cryo-TEM, both methods produce very similar size distributions for each of the LNP formulations. However, the shape of the R_h distributions observed by DLS, while extending over the same approximate R_h range determined by AUC and cryo-TEM, appears to exaggerate the width and overemphasize larger components of the R_h distributions (see Figure 4.4). This can be attributed to the different scaling in DLS measurements, which scale with the sixth power of the particle radius, and the fact that DLS suffers from significantly lower resolution and sensitivity for heterogeneity than AUC or cryo-TEM. This suggests that DLS has limited utility for characterizing LNP formulations compared to MW-AUC.

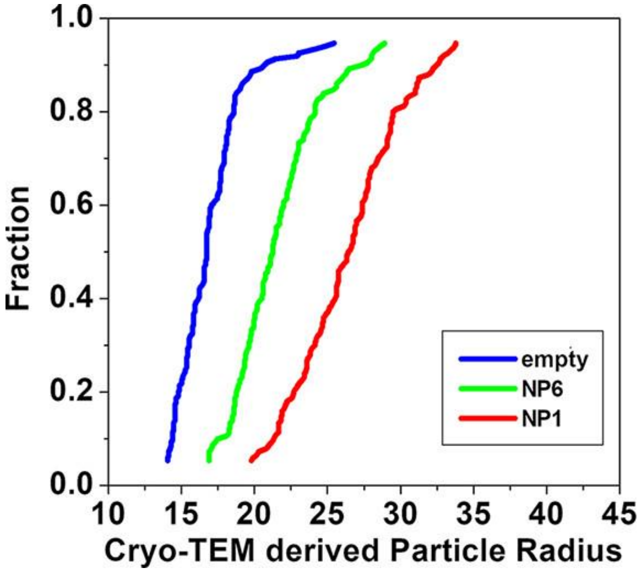


Figure 4.3: Cryo-TEM derived particle radii for empty LNP (blue), NP6 (green), and NP1 (red) LNPs. Total number of particles counted: 150.

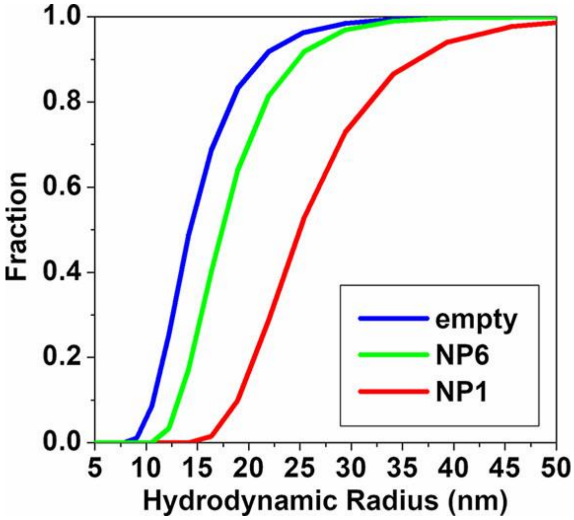


Figure 4.4: Number-average corrected hydrodynamic radii as measured by dynamic light scattering for empty (blue), NP6 (green), and NP1 (red) LNP preparations.

4.5.3 Analysis of Multi-wavelength AUC Data

The analysis of the MW-AUC data shows that sedimentation profiles derived from the siRNA signal largely coincide with the sedimentation profiles derived from the LNP signal. This indicates that all siRNA molecules are fully complexed with lipids, in full agreement with the PSV analysis, which also did not detect siRNA free in solution. It also suggests that the NP1 and NP6 formulations do not contain empty LNPs. This was determined by spectrally separating the sedimentation distributions for the lipid and siRNA cargo components, making it possible to follow each component separately and to compare the sedimentation profiles of the spectrally separated species (LNP and siRNA in this case) [14, 15]. If free siRNA were present in the formulation, sedimentation profiles would diverge significantly and display positive sedimentation signals even in higher D_2O concentrations because of the significantly lower PSV of free nucleic acid (compare Figure 4.1). The MW-AUC data for NP1 are shown in Figure 4.5 and indicate that the siRNA sedimentation profile closely follows the sedimentation profile of the lipid component. The sedimentation profiles indicate a slight bias toward faster sedimentation for the lipid component over the siRNA component. It should be noted that this deviation in the sedimentation rate between the two signals is likely due to the fact that the LNP lipid shell does not absorb light, but any apparent absorbance is the result of Mie scattering. Hence, the scaling of the LNP signal is proportional to the sixth power of the shell’s radius, while the siRNA absorbance is proportional to the siRNA mass, which scales only with the third power of the shell’s radius. The different scalings would therefore overemphasize the larger LNP particles in the LNP signal, and small, lipid bound siRNA samples would not be visible in the LNP signal and instead be enhanced for very small particles in the spectrally separated siRNA component. This effect is most obvious with the largest LNP structures, which we observed primarily with the NP1 samples. A brief overview of the multi-wavelength AUC approach is presented in Section SF 4.2.

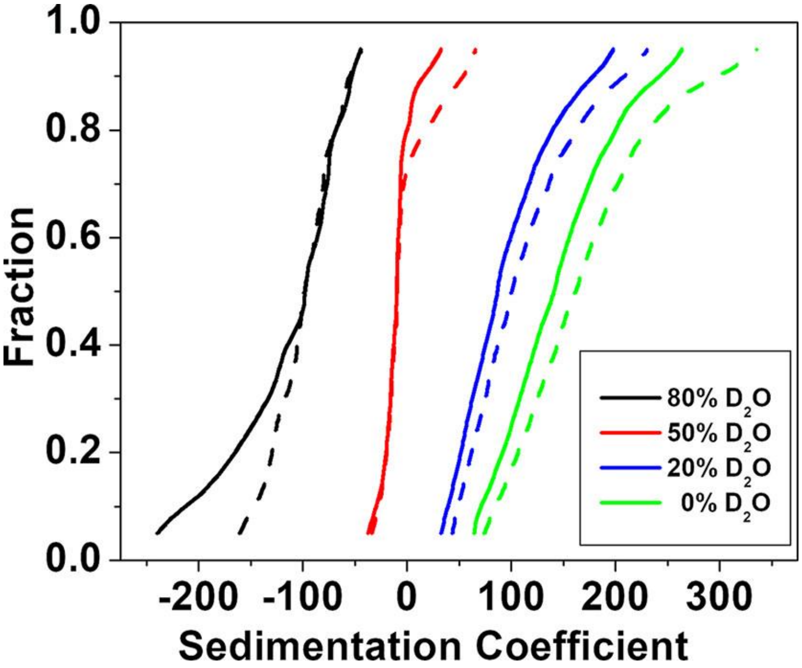


Figure 4.5: MW-AUC sedimentation coefficient distributions for NP1 in four different D_2O concentrations. Dashed lines represent the light-scattering signal from the lipid component, while solid lines represent the siRNA absorbance.

4.5.4 Validation of siRNA Incorporation in LNPs

To further validate the complete incorporation of siRNA into the LNP, we performed a separate control experiment where we measured free siRNA in the same light water buffer used for NP1 and NP6, observing a homogeneous distribution sedimenting at 2.58 s, which is in stark contrast to the sedimentation coefficient distribution observed for NP1 (ranging between ~50 and ~250 s). In a third orthogonal validation, we separately prepared NP1, formulated with fluorescently labeled siRNA, and measured it in the same light water buffer used for the unlabeled NP1 sample in an analytical ultracentrifuge equipped with a fluorescence optical system. In this experiment, only the fluorescently labeled siRNA will be detectable. The SVEs for these three experiments were analyzed by the method of van Holde–Weischet [11], using integral sedimentation coefficient distributions. The observed distribution of the fluorescence experiment exactly matched the sedimentation distribution from the measurement of NP1 loaded with unlabeled siRNA, measured at 260 nm, proving that all siRNA must be incorporated in LNPs as predicted by the PSV and multi-wavelength experiments. These results are summarized in Figure 4.6.

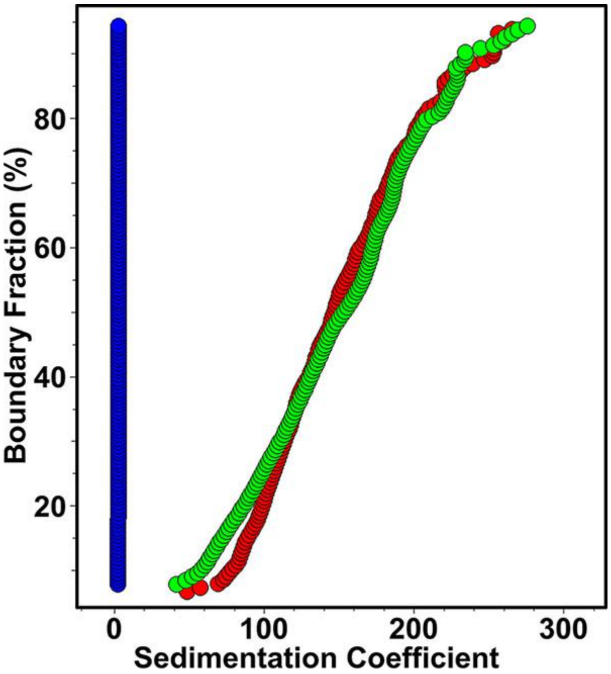


Figure 4.6: Validation of siRNA loading by van Holde–Weischet integral sedimentation distribution comparison between NP1 measured at 260 nm absorbance (red) and NP1 loaded with fluorescently labeled siRNA, detected by fluorescence emission (green), and pure siRNA

4.6 Conclusions

We have shown that three orthogonal experimental approaches based on analytical ultracentrifugation (density matching and multi-wavelength and fluorescence emission) can be used to provide high-resolution composition information for LNP formulations and characterize LNP loading with nucleic acids. These techniques excel in terms of resolution and information content when compared to traditional methods such as light scattering or cryo-TEM, providing detailed information about LNP loading with nucleic acids. In order to distinguish loaded from empty LNPs we exploit (a) the difference in density between lipids and nucleic acids and (b) the difference in absorbance profiles between profiles from empty LNPs and from nucleic acids, and validated our results with fluorescence measurements. Our results show that density matching and MW-AUC can be used in tandem and extend the utility of analytical ultracentrifugation to provide reliable distributions for mass, size, and partial specific volume. We conclude that this approach represents an important alternative for the analysis of LNP formulations and can be used for routine validation of LNP formulations used for important therapeutic applications such as vaccines and gene therapy.

4.7 Methods and Experiments

4.7.1 Preparation and Analysis of Lipid Nanoparticles

LNP formulations were prepared as previously described [19]. Briefly, lipid components such as ionizable cationic lipid (KC2), distearoylphosphatidylcholine (DSPC), cholesterol, and PEG lipid were dissolved in ethanol at a ratio of 50/10/38.5/1.5 (mol %), respectively. The ethanol phase was combined with the nucleic acid aqueous phase (at pH 4) through a T-junction mixer [24, 25] at a total flow rate of 20 mL/min. The resulting suspension was dialyzed against phosphate buffered saline (pH 7.4) overnight. Before AUC analysis, LNP concentrations were adjusted to range between 0.1 and 1.0 OD in the measured wavelength range. Encapsulation was determined using the RiboGreen assay [26]. Lipid content was measured using a Total Cholesterol assay kit (Wako Diagnostics, Mountain View, CA, USA), and total lipid concentration was extrapolated. LNPs were loaded with double-stranded, 21 base-pair methylated siRNA against firefly luciferase, purchased from Integrated DNA Technologies (IDT, Coralville, IA, USA) [27]. Fluorescently labeled siRNA was generated by incubating RiboGreen (Invitrogen, Carlsbad, CA, USA) with siRNA prior to incorporation into LNP. The labeling ratio used was 48uL RiboGreen reagent per 1 mg of siRNA. Uncomplexed RiboGreen was removed during the overnight dialysis.

4.7.2 Cryogenic Transmission Electron Microscopy

Cryo-TEM was performed as previously described [19]. A 3–5 μL drop of concentrated LNP solution was added to a glow-discharged TEM grid. A Mark IV Vitrobot (FEI, Hillsboro, OR, USA) was used to vitrify samples in liquid ethane. The frozen samples were maintained under liquid nitrogen until imaged. Imaging was performed using a FEI Tecnai LaB6 G2 TEM operating at 200 kV with a bottom-mount CCD Eagle 4k detector (FEI). Alternatively, imaging was performed on an FEI Titan Krios operating at 300 kV with a Falcon III direct electron detector. A nominal underfocus of 1 μm was used to enhance contrast. LNPs were prepared using standard rapid-mixing methods [19], dialyzed into neutral buffer, and concentrated to achieve 15–30 mg/mL total lipid.

4.7.3 Analytical Ultracentrifugation

All AUC experiments were performed at the Canadian Center for Hydrodynamics at the University of Lethbridge, Alberta, Canada using an Optima AUC (Beckman Coulter). LNPs loaded with fluorescently labeled siRNA were measured in a Proteomelab XL-A (Beckman-Coulter) equipped with an Aviv fluorescence detector (excitation wavelength, 488 nm) using 3 mm titanium fluorescence assemblies (Nanolytics Instruments, Germany). All samples were measured in an An60Ti rotor, and in phosphate buffered saline at 20 °C. Samples measured using UV-intensity optics were loaded into Epon-charcoal centerpieces and fitted with quartz windows. siRNA and plasmid DNA were measured at 260 nm, siRNA was measured at 60 krpm. and 500 scans were analyzed. Plasmid DNA was measured at 40 krpm, and 200 scans were analyzed. BSA was measured at 280 nm and 45 krpm, and 160 scans were analyzed. For LNP preparations, the rotor speed was set so that each LNP preparation provided sufficient sedimentation signal, with at least 20 scans collected per wavelength. This speed varied between D_2O concentrations and LNP preparations, and ranged between 12 and 14 krpm. All Optima AUC experiments were performed in intensity mode and analyzed with UltraScan-III, rel. 5843, as described in ref [6]. MW-AUC experiments were measured between 230 and 290 nm, in 2 nm increments. SVEs are analyzed by fitting experimental data to linear combinations of finite element solutions of the Lamm equation [28] using the two-dimensional spectrum analysis (2DSA), which provides concentration distributions for the sedimentation and diffusion coefficients of each solute in a mixture [7]. LNPs are a convenient target for measuring their loading efficiency by SVE because their PSVs vary as a function of drug encapsulation due to the fact that nucleic acids are relatively dense compared to the hydrated shell (lipids and hydrated PEG molecules), resulting in a sedimenting particle that has a density close to that of light water. We describe here how the variation in PSV can be exploited by D_2O density matching experiments, and a global analysis of SVEs can be used to derive distributions that show this variation in PSV and allow us to distinguish the LNP loading state. We recall that a solute's sedimentation speed is a function of its PSV, as represented by the Svedberg relationship, shown in (Equation 4.1:

$$s = \frac{M(1 - \bar{v}p)}{N_A f}$$

(Equation 4.1)

where s is the sedimentation coefficient, M is the molar mass, N_A is Avogadro's number, f is the frictional coefficient, ρ is the density of the solvent, and $\bar{v}$ is the PSV of the analyte. In cases where the inverse of the buffer density equals the PSV of the particle, the buoyancy term, $1 - \bar{v}\rho$, vanishes, and no sedimentation will be observed. Therefore, by modulating the density of the solvent, it is possible to alter the sedimentation behavior of each molecule in a mixture by an amount proportional to its PSV. This effect can be exploited with a density matching experiment [17]. In this type of experiment, the density (and viscosity) of an aqueous buffer is precisely modulated by preparing the buffer with different ratios of heavy and light water, while preserving the ionic strength of the buffer.

When an analyte is sedimented in buffers with varying densities, the observed change in the sedimentation coefficient can be plotted as a function of buffer density and then extrapolated to obtain the density at $s = 0$, where $1/\rho = \bar{v}$ [29, 30]. When performing the experiment in a range of light and heavy water ratios, each molecule in the mixture with a different PSV will produce a different extrapolation. We exploit this phenomenon to characterize the PSV composition of a heterogeneous mixture of molecules with varying PSVs. Here, we demonstrate that replicate SVEs of LNPs, performed in buffers with different H_2O:D_2O ratios, can be used to derive the LNP loading state. Using the additional information on their spherical shape from cryo-TEM experiments, we can further derive the molar mass and hydrodynamic radius.

By performing the experiment in multi-wavelength mode (scanning between 230 and 290 nm in 2 nm increments), it is possible to optically resolve the contributions from the lipids and the nucleic acids, since they differ in their spectral contributions over the examined spectral range (detailed descriptions of the MW-AUC method can be found in the literature; [14-16, 31] an introductory explanation of the method is provided in SF 4.2). Nucleic acids produce a characteristic 260 nm absorbance band, while the lipid nanoparticles do not absorb at all, but instead display a characteristic Mie-scattering profile with a monotonically increasing absorbance signal with decreasing wavelength (see Figure 4.7). Capitalizing on these spectral differences, it is possible to resolve siRNA signal from lipid signal by using a non-negatively constrained least-squares decomposition of the wavelength scans obtained in the multi-wavelength experiment and analyze their hydrodynamic properties separately. By examining sedimentation coefficient distributions for each signal separately, it is possible to determine if the two molecules co-sediment as a single particle with identical sedimentation coefficients, in which case the siRNA is encapsulated by the liposome, or if the RNA is free in solution. In the latter case, their sedimentation and PSV distributions would be significantly different because of the buoyancy differences. We have developed an algorithm capable of combining all analysis steps required to deduce distributions for the partial specific volume, and implemented it in a module (termed *us_buoyancy*) that is now part of the UltraScan software suite [20]. UltraScan is a multiplatform, open-source software that can be freely downloaded from our Web site (http://ultrascan3.aucsolutions.com). Our approach employs the following steps:

1. For a density matching experiment, an equal amount of LNP sample is diluted into phosphate buffered saline, pH 7.4, prepared in four different ratios of H_2O:D_2O ranging between 0% and 99% D_2O. In order to achieve conditions with high D_2O content, it is recommended to prepare buffers in D_2O directly and to dilute a highly concentrated LNP sample with the D_2O buffer to minimize the amount of light water in the final sample. It is important to maintain the same analyte concentration in each replicate experiment to avoid oligomerization effects, which could alter the sedimentation coefficient distribution. To obtain a 20× D_2O buffer stock solution, we prepared a 20× buffer in light water, dried an aliquot in a lyophilizer, and resuspended the pellet in D_2O to create 20× D_2O buffer. An assumption is made that the LNPs do not change chemically in response to variable D_2O concentrations and maintain their assembly state composition. Whenever D_2O is employed, the potential for HDX exists, which affects the

absolute PSV measured with variable effect on different macromolecules, with nucleic acids and proteins more sensitive to HDX than aliphatic groups in lipids [22, 32]. Investigators concerned with the impact of HDX on the PSV can substitute D_2O with the significantly more expensive $H_2{}^{18}O$, which has physical properties similar to those of D_2O, but avoids deuterium exchange artifacts.

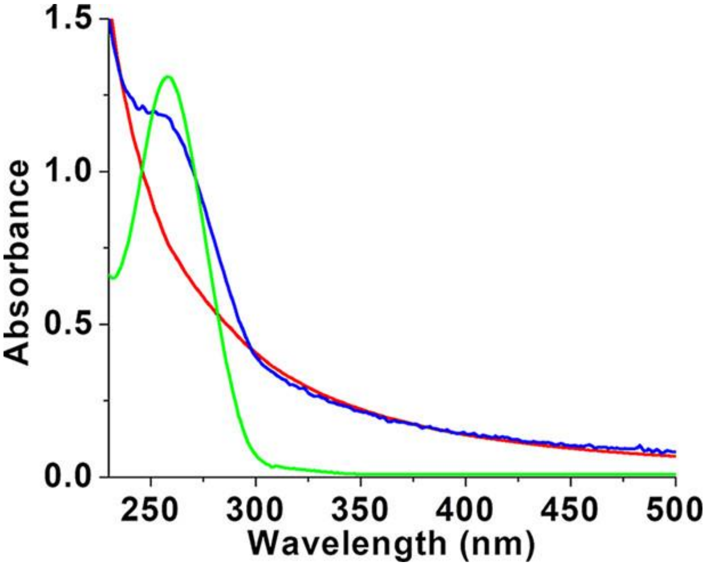


Figure 4.7: UV spectral properties of RNA (green), empty LNP (red), and RNA-loaded LNPs (blue).

2. In general, it is preferable to perform the sedimentation velocity experiments at low enough sample concentrations to maintain near ideal transport behavior but high enough to still produce sufficient signal across the examined wavelength range, utilizing the full dynamic range of the detector. Rotor speeds should be fast enough to optimize sedimentation separation of solutes, while at the same time slow enough to obtain sufficient scan numbers. For multi-wavelength sedimentation velocity experiments in the Optima AUC instrument, it is recommended to use a speed that synchronizes the lamp's flash rate with the rotor speed; this will minimize scanning time and allow a sufficient number of scans to be recorded for each wavelength before the sample is pelleted. For LNPs investigated at the Canadian Center for

Hydrodynamics this condition is matched at speeds between 12,000 and 14,000 rpm, which results in a 19–17 s scan time, respectively, for both cell channels.

3. Each experimental data set is analyzed according to standard workflows implemented in UltraScan to derive an iterative two-dimensional spectrum analysis (2DSA) model for each triple [7], which is defined to be the data resulting from one cell, one channel, and a single wavelength. Our approach includes a stepwise refinement of the fit by first removing time-invariant noise contributions from the data and then adding removal of radially-invariant noise contributions, and to fit the meniscus and cell bottom positions. Finally, an iterative refinement of the two-dimensional parameter space is performed. Additional refinement by the 2DSA-Monte Carlo analysis [8] or the parametrically constrained spectrum analysis [5] is recommended when additional regularization is needed. The 2DSA method requires the generation of a grid that spans the parameter space of the sedimentation and diffusion coefficients describing the experimental system. This grid typically uses a parametrized grid based on the sedimentation coefficient and the frictional ratio, φ, describing the anisotropy of each species [1]. The parametrization of the diffusion coefficient using φ requires an initial guess of the value of the PSV, which is typically assumed to be constant for all species to be fitted, or bimodal for experiments which contain both sedimenting and floating solutes. It is important to note that even if the actual values of the PSV differ, the grid can nevertheless accurately represent the appropriate sedimentation and diffusion coefficients, as long as the frictional ratio φ covers all sedimentation and diffusion coefficients needed to explain the experimental data. Whether this condition is met can be assessed by the RMSD and the randomness of the residuals produced in the fit. Hence, accurate values for the PSV are not needed *a priori* to obtain a satisfactory fit of the experimental data. In UltraScan, conversion of observed sedimentation and

diffusion coefficients to standard conditions (based on buffer viscosity, density, and experimental temperature) is automatically performed, producing models composed of multiple discrete species, where each species is represented by a $s_{20,W}$ value, a $D_{20,W}$ value, and a partial concentration. For this application it is important that the density increment due to D_2O is not added to the buffer correction, since the effect of the density variation contributed by D_2O is used to globally obtain PSV distributions. However, viscosity corrections due to the addition of D_2O must be included. Examples for such plots obtained for LNP preparations containing two different loading ratios are shown in Figure 4.8.

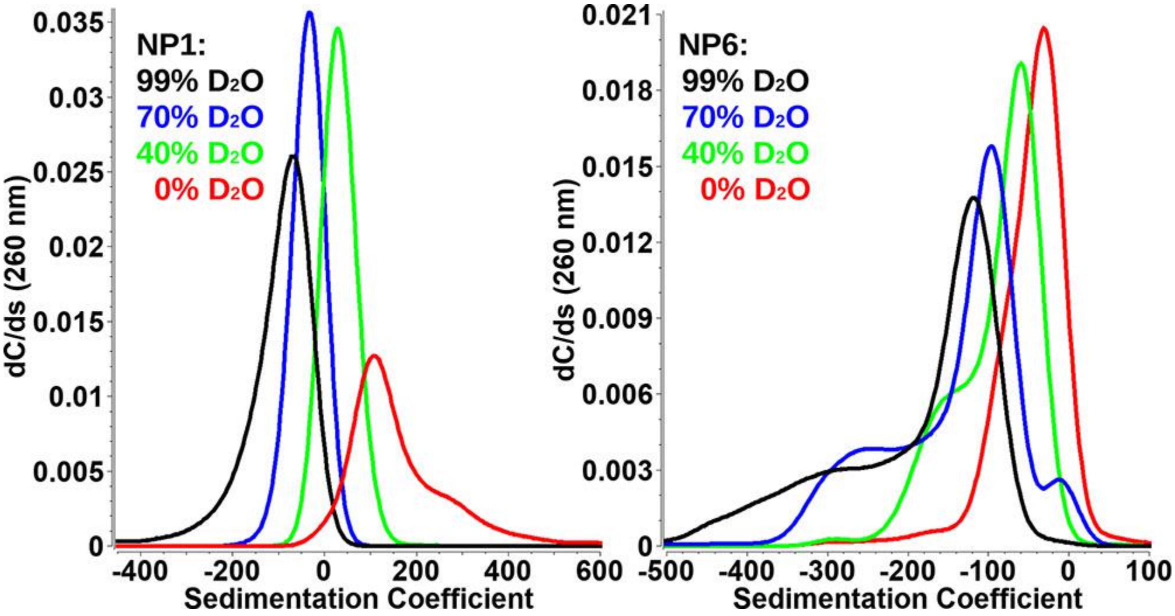


Figure 4.8: Sedimentation coefficient distributions for density matching SVEs of two different LNP loading ratios (phospholipid:RNA backbone phosphate). NP1, containing one phospholipid molecule per siRNA phosphate is shown on the left; NP6, containing six phospholipid molecules per siRNA phosphate is shown on the right. Distributions are shown for four buffers containing 0–99% D_2O (0%, red; 40%, green; 70%, blue; 99%, black).

4. Integral distributions for sedimentation and diffusion coefficients, based on the hydrodynamically corrected values, are derived from the final model distributions obtained in item 3, where the diffusion coefficient (*D*) is given by:

$$D = \frac{RT}{Nf}$$

(Equation 4.2)

where R is the universal gas constant and T is the temperature in kelvin. Like s, D is inversely proportional to the frictional properties of the molecule. Diffusion distributions ($D_{20,\mathrm{W}}$) from all D_2O concentrations can be averaged across measurements from all buffers but only for values taken from identical boundary fractions. Since diffusion is only affected by viscosity, a value that is already corrected, no further correction needs to be applied to the diffusion values measured from different buffers. Also, since the analyte concentration is held constant, equivalent boundary fractions correspond to the same sedimenting species across all buffer densities. Since the analyte concentration in each D_2O concentration sample is identical, integral distributions will reflect the same sample concentration and can therefore be normalized to account for minor variations in cell pathlengths. Examples of integral sedimentation coefficient distributions for NP6 LNPs in different H_2O:D_2O mixtures are shown in Figure 4.9. Boundary fractions near the extremes (5–10%) are sensitive to experimental noise and therefore tend to be less reliable and are typically excluded from the plots.

5. For each corresponding boundary fraction from each buffer, the corresponding viscosity-corrected s values and $D_{20,\mathrm{W}}$ values are recorded. The s values are now plotted against the density of each buffer and extrapolated to the zero sedimentation point. The density of the solution at the zero sedimentation point represents the density where the species in this boundary fraction is exactly buoyant in the buffer solution. The inverse of this density equals the PSV of the sample, and since each boundary fraction can extrapolate to a different position, this procedure generates

a unique partial specific volume for each boundary fraction. The collection of all extrapolated boundary fractions represents the PSV distribution for the mixture.

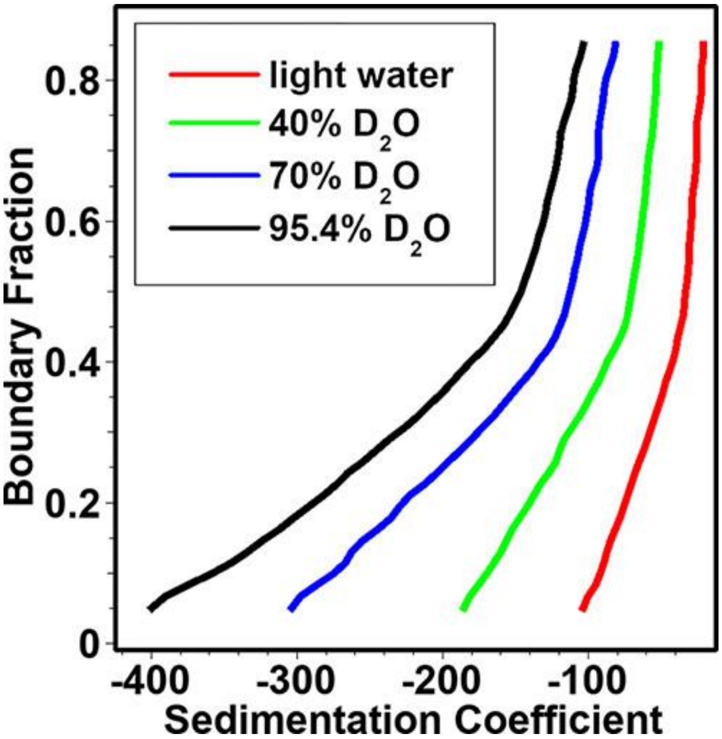


Figure 4.9: Integral sedimentation coefficient distributions for NP6 in four different D2O concentrations (0%, red; 40%, green; 70%, blue; 95.4%, black).

6. Using the boundary fraction arrays for the PSV, and the $s_{20,\mathrm{W}}$ and $D_{20,\mathrm{W}}$ values obtained in light water, additional properties of the sample composition can now be derived numerically. Combining and rearranging (Equation 4.1 and (Equation 4.2 directly yield the molar mass:

$$M = \frac{sRT}{D(1 - \bar{v}p)}$$

(Equation 4.3)

The frictional coefficient of each particle can be obtained from the diffusion coefficient:

$$f = \frac{RT}{ND}$$

(Equation 4.4)

Combining this result with the Stokes–Einstein relationship, the hydrodynamic radius, R_h, can be obtained:

$$R_h = \frac{f}{6\pi\eta}$$

(Equation 4.5)

where η signifies the viscosity of the solvent. Representing the volume of each particle as a perfect sphere, the minimal radius, R_0, of each particle (Equation 4.6) and the corresponding frictional coefficient, f_0, can be derived from the Stokes–Einstein relationship (Equation 4.7):

$$R_0 = \left(\frac{3M\bar{v}}{4\pi N}\right)^{\frac{1}{3}}$$

(Equation 4.6)

$$f_0 = 6\pi\eta R_0$$

(Equation 4.7)

Finally, the anisotropy, φ, or frictional ratio, of the particle can be obtained:

$$\varphi = \frac{f}{f_0}$$

(Equation 4.8)

Screenshots of the UltraScan module dialogues used to generate distributions from these parameters and display them as integral distribution graphs are shown in Supplemental Information SF 4.4 and 4.5.

In some cases, additional constraints can be applied [4]. For LNPs, cryo-TEM results suggest the presence of mostly spherical particles (see Figure 4.10). If particles are spherical as confirmed by a separate technique such as cryo-TEM, the assumption can be made that φ (equation 4.8) for most particles corresponds to an anisotropy of unity. This is helpful in cases where the diffusion signal is weak, either because the sedimentation is too fast, not allowing enough time for sufficient diffusion signal to be collected, or the heterogeneity is so large that the signal of the diffusion coefficient for individual species is simply too small to be of any value. In the case of LNPs examined here, both conditions apply, and an additional constraint based on the anisotropy allows us to calculate the appropriate diffusion coefficient by transforming the sedimentation and PSV data pairs with equation 4.9, using φ = 1.0:

$$D = \frac{RT}{9\pi\eta\varphi N}\left(\frac{2s\varphi\bar{v}\eta}{1-\bar{v}p}\right)^{-0.5}$$

(Equation 4.9)

This constraint can be incorporated into the fitting equation and tested for validity by confirming random residuals for the fitted data. To further validate our assumption of φ = 1.0, we offer the following observations: Even though the actual hydration layer, including PEG molecules, cannot be observed by cryo-TEM, it is reasonable to expect that the hydration layer does not substantially alter the morphology of LNPs and that it simply replicates the spherical shape observed in the cryo-TEM images (see Figure 4.10). Furthermore, simulations with φ values larger than unity yield hydrodynamic radii no longer consistent with the DLS measurements of the hydrodynamic radius. We show an example of predictions for hydrodynamic radius and molar mass based on higher anisotropies ($\varphi = 1.2$, $\varphi = 1.5$) in SF 4.6.

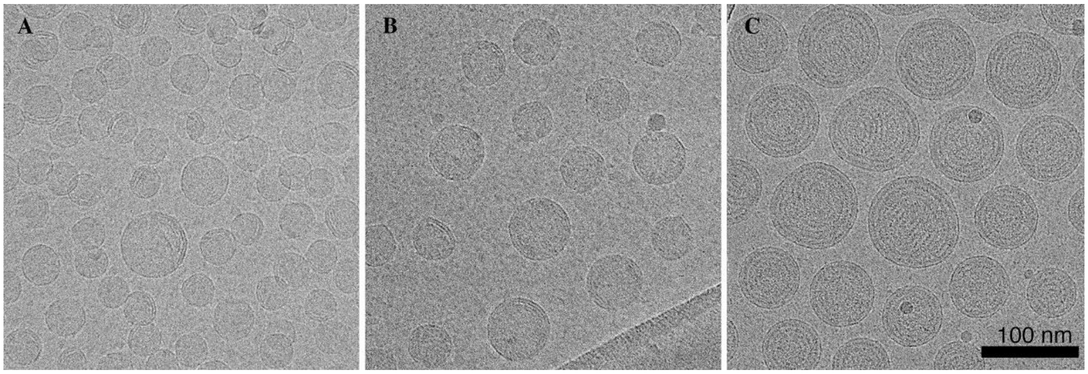


Figure 4.10: Cryo-TEM images of empty LNP (A), NP6 filled LNPs (B), and NP1 filled LNPs (C) suggesting the presence of mostly spherical particles, regardless of loading stage, justifying the use of the anisotropy constraint $\varphi = 1.0$.

4.8 Limitations

We noticed, that for heterogeneous samples, the density matching algorithm is very sensitive to the absorbance baseline determined while deriving the time-invariant noise during fitting [33]. When baselines did not precisely align between SVEs performed in different buffers, any larger steps in partial specific volume for different molecular mixtures, for example, mixtures of non-interacting proteins and DNA, can cause slight shifts in boundary fractions, no longer guaranteeing a precise alignment of solutes represented in specific boundary fractions. This can cause erratic extrapolations in the transition region of the boundary fractions, which can be avoided by lowering the number of divisions along the boundary. Also, the problem can be minimized by carefully fitting the original SVEs, making sure that any buffer absorbance is properly accounted for in the time-invariant noise component, leading to a zero baseline for all data sets included in the density matching extrapolation.

4.9 References

1. Kim H, Brookes E, Cao W, Demeler B. Two-dimensional grid optimization for sedimentation velocity analysis in the analytical ultracentrifuge. Eur Biophys J. 2018;47(7):837-844. doi: 10.1007/s00249-018-1309-z.

2. Williams TL, Gorbet GE, Demeler B. Multi-speed sedimentation velocity simulations with UltraScan-III. Eur Biophys J. 2018;47(7):815-823. doi: 10.1007/s00249-018-1308-0.

3. Gorbet GE, Mohapatra S, Demeler B. Multi-speed sedimentation velocity implementation in UltraScan-III. Eur Biophys J. 2018;47(7):825-835. doi: 10.1007/s00249-018-1297-z.

4. Demeler B, Nguyen T-L, Gorbet GE, Schirf V, Brookes EH, Mulvaney P, et al. Characterization of Size, Anisotropy, and Density Heterogeneity of Nanoparticles by Sedimentation Velocity. Anal Chem. 2014;86(15):7688-7695. doi: 10.1021/ac501722r.

5. Gorbet G, Devlin T, Hernandez Uribe Blanca I, Demeler Aysha K, Lindsey Zachary L, Ganji S, et al. A Parametrically Constrained Optimization Method for Fitting Sedimentation Velocity Experiments. Biophys J. 2014;106(8):1741-1750. doi: https://doi.org/10.1016/j.bpj.2014.02.022.

6. Demeler B. Methods for the design and analysis of sedimentation velocity and sedimentation equilibrium experiments with proteins. Current protocols in protein science. 2010;60(1):7.13. 11-17.13. 24.

7. Brookes E, Cao W, Demeler B. A two-dimensional spectrum analysis for sedimentation velocity experiments of mixtures with heterogeneity in molecular weight and shape. Eur Biophys J. 2010;39(3):404-414.

8. Demeler B, Brookes E. Monte Carlo analysis of sedimentation experiments. Colloid and Polymer Science. 2008;286:129-137.

9. Brookes E, Demeler B. Parallel computational techniques for the analysis of sedimentation velocity experiments in UltraScan. Colloid and Polymer Science. 2008;286:139-148.

10. Brookes E, Demeler B. Parsimonious Regularization Using Genetic Algorithms Applied to the Analysis of Analytical Ultracentrifugation Experiments. GECCO ’07: Proceedings of the 9th Annual Conference on Genetic and Evolutionary Computation, vols. 361–368, July 2007. ACM 978-1-59593-697-4/07/0007. 2007. doi: https://doi.org/10.1145/1276958.1277035.

11. Demeler B, Van Holde KE. Sedimentation velocity analysis of highly heterogeneous systems. Anal Biochem. 2004;335(2):279-288.

12. Cao W, Demeler B. Modeling analytical ultracentrifugation experiments with an adaptive space-time finite element solution for multicomponent reacting systems. Biophys J. 2008;95(1):54-65. Epub 2008/04/09. doi: 10.1529/biophysj.107.123950. PubMed PMID: 18390609; PubMed Central PMCID: PMCPMC2426643.

13. Cao W, Demeler B. Modeling analytical ultracentrifugation experiments with an adaptive space-time finite element solution of the Lamm equation. Biophys J. 2005;89(3):1589-1602. Epub 2005/06/28. doi: 10.1529/biophysj.105.061135. PubMed PMID: 15980162; PubMed Central PMCID: PMCPMC1366663.

14. Zhang J, Pearson JZ, Gorbet GE, Cölfen H, Germann MW, Brinton MA, et al. Spectral and hydrodynamic analysis of West Nile Virus RNA–protein interactions by multiwavelength sedimentation velocity in the analytical ultracentrifuge. Anal Chem. 2017;89(1):862-870.

15. Gorbet GE, Pearson JZ, Demeler AK, Cölfen H, Demeler B. Next-generation AUC: analysis of multiwavelength analytical ultracentrifugation data. Methods Enzymol. 562: Elsevier; 2015. p. 27-47.

16. Pearson JZ, Krause F, Haffke D, Demeler B, Schilling K, Cölfen H. Chapter One - Next-Generation AUC Adds a Spectral Dimension: Development of Multiwavelength Detectors for the Analytical Ultracentrifuge. In: Cole JL, editor. Methods Enzymol. 562: Academic Press; 2015. p. 1-26.

17. Swygert SG, Manning BJ, Senapati S, Kaur P, Lindsay S, Demeler B, et al. Solution-state conformation and stoichiometry of yeast Sir3 heterochromatin fibres. Nature communications. 2014;5(1):1-12.

18. Kulkarni JA, Witzigmann D, Chen S, Cullis PR, van der Meel R. Lipid Nanoparticle Technology for Clinical Translation of siRNA Therapeutics. Accounts of Chemical Research. 2019;52(9):2435-2444. doi: 10.1021/acs.accounts.9b00368.

19. Kulkarni JA, Darjuan MM, Mercer JE, Chen S, van der Meel R, Thewalt JL, et al. On the Formation and Morphology of Lipid Nanoparticles Containing Ionizable Cationic Lipids and siRNA. ACS Nano. 2018;12(5):4787-4795. doi: 10.1021/acsnano.8b01516.

20. Demeler B, Gorbet GE. In Analytical Ultracentrifugation: Instrumentation, Software, and Applications (eds Uchiyama S., Arisaka F., Stafford W. F. & Laue T.) 119-143 (Springer Japan). 2016. doi: DOI: 10.1007/978-4-431-55985-6_8.

21. Largy E, Gabelica V. Native Hydrogen/Deuterium Exchange Mass Spectrometry of Structured DNA Oligonucleotides. Anal Chem. 2020;92(6):4402-4410. doi: 10.1021/acs.analchem.9b05298.

22. Murata K, Wolf M. Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. Biochimica et Biophysica Acta (BBA) - General Subjects. 2018;1862(2):324-334. doi: https://doi.org/10.1016/j.bbagen.2017.07.020. PubMed PMID: MURATA2018324.

23. Franken LE, Boekema EJ, Stuart MCA. Transmission Electron Microscopy as a Tool for the Characterization of Soft Materials: Application and Interpretation. Adv Sci (Weinh). 2017;4(5):1600476. Epub 2017/05/27. doi: 10.1002/advs.201600476. PubMed PMID: 28546914; PubMed Central PMCID: PMCPMC5441488.

24. Jeffs LB, Palmer LR, Ambegia EG, Giesbrecht C, Ewanick S, MacLachlan I. A Scalable, Extrusion-Free Method for Efficient Liposomal Encapsulation of Plasmid DNA. Pharm Res. 2005;22(3):362-372. doi: 10.1007/s11095-004-1873-z.

25. Kulkarni JA, Tam YYC, Chen S, Tam YK, Zaifman J, Cullis PR, et al. Rapid synthesis of lipid nanoparticles containing hydrophobic inorganic nanoparticles. Nanoscale. 2017;9(36):13600-13609. doi: 10.1039/C7NR03272B.

26. Chen S, Tam YYC, Lin PJC, Leung AKK, Tam YK, Cullis PR. Development of lipid nanoparticle formulations of siRNA for hepatocyte gene silencing following subcutaneous administration. J Controlled Release. 2014;196:106-112. doi: https://doi.org/10.1016/j.jconrel.2014.09.025.

27. Basha G, Ordobadi M, Scott WR, Cottle A, Liu Y, Wang H, et al. Lipid Nanoparticle Delivery of siRNA to Osteocytes Leads to Effective Silencing of SOST and Inhibition of Sclerostin In Vivo. Molecular Therapy - Nucleic Acids. 2016;5:e363. doi: https://doi.org/10.1038/mtna.2016.68.

28. Lamm O. Die differentialgleichung der ultrazentrifugierung: Almqvist & Wiksell; 1929.

29. Edelstein SJ, Schachman HK. Measurement of partial specific volume by sedimentation equilibrium in H2O・D2O solutions. Methods Enzymol. 27: Academic Press; 1973. p. 82-98.

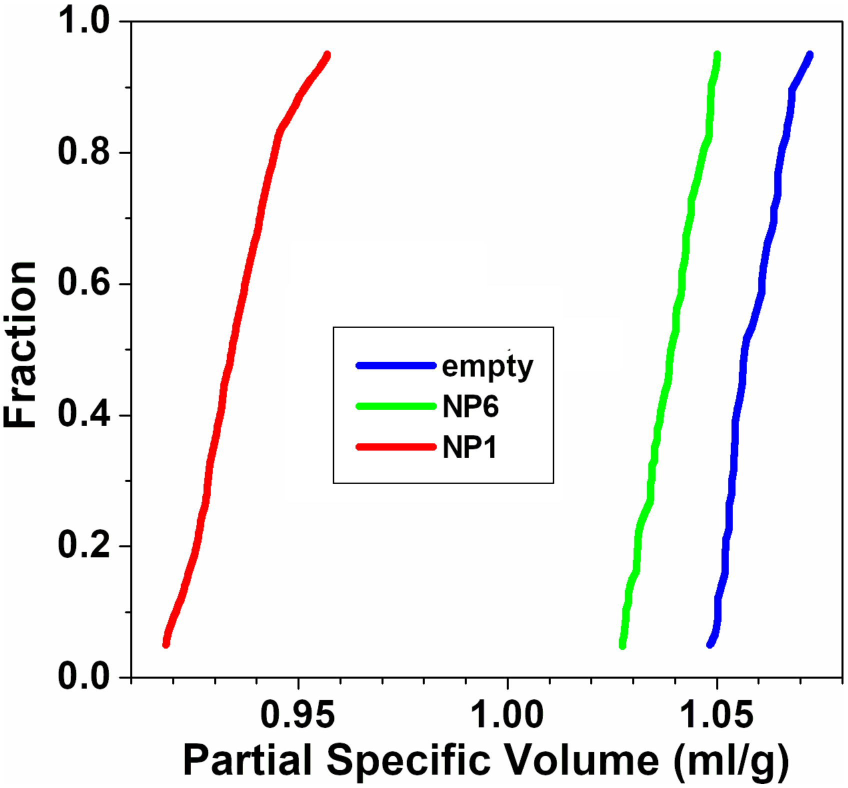
30. Gohon Y, Pavlov G, Timmins P, Tribet C, Popot J-L, Ebel C. Partial specific volume and solvent interactions of amphipol A8-35. Anal Biochem. 2004;334(2):318-334. doi: https://doi.org/10.1016/j.ab.2004.07.033.

31. Demeler B. Measuring molecular interactions in solution using multi-wavelength analytical ultracentrifugation: combining spectral analysis with hydrodynamics. Biochem (Lond). 2019;41(2):14-18. Epub 2019/01/01. doi: 10.1042/bio04102014. PubMed PMID: 32952314; PubMed Central PMCID: PMCPMC7500494.

32. Breyton C, Gabel F, Lethier M, Flayhan A, Durand G, Jault J-M, et al. Small angle neutron scattering for the study of solubilised membrane proteins. The European Physical Journal E. 2013;36(7):71. doi: 10.1140/epje/i2013-13071-6.

33. Schuck P, Demeler B. Direct Sedimentation Analysis of Interference Optical Data in Analytical Ultracentrifugation. Biophys J. 1999;76(4):2288-2296. doi: https://doi.org/10.1016/S0006-3495(99)77384-4.

4.10 Supplemental Information



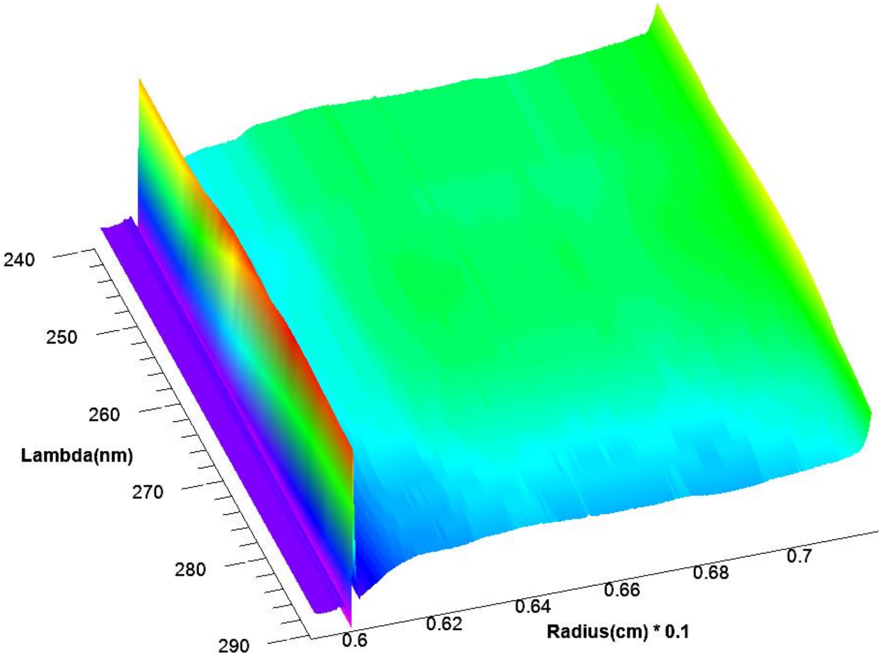
SF 4.1: Expanded view of the partial specific volume plot shown in Figure 9 to emphasize the observed heterogeneity of PSV

SI 4.1: Multi-wavelength Analytical Ultracentrifugation (MW-AUC):

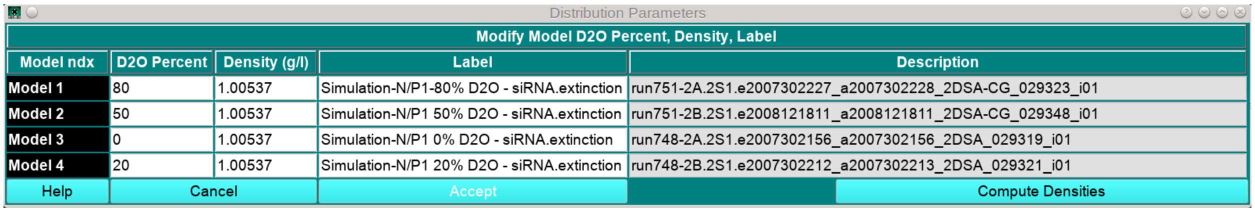
MW-AUC is a recent technique enabled by new instruments that allows the user to collect sedimentation velocity data in intensity mode at multiple wavelengths. This is useful when the sample includes multiple analytes, each with a different and distinct chromophore. In this study, the different chromophores derive from siRNA and from the Mie scattering of LNPs in the UV. When the spectra of each analyte are known *a priori*, it is possible to deconvolute the contribution of each analyte to each hydrodynamically distinct species in the mixture. In this case this allows us to determine if LNPs and siRNA sediment as free, individual species, or as a complex. If molar extinction coefficients are known for each measured wavelength it is possible to assign molar quantities of each analyte to each hydrodynamic species, otherwise, relative quantities can easily be determined from the spectral deconvolution. This approach has wide-reaching applications for studies of interacting or complex-forming systems as is present in this study. It provides a second, spectral dimension to aid in the characterization of the solutes present in a mixture, in addition to the hydrodynamic dimension, which separates individual species based on their hydrodynamic properties.

Approach:

Sedimentation velocity experiments (SVEs) for each wavelength are initially analyzed according to standard analysis protocols as described in [1]. In the last refinement step, an iterative two-dimensional spectrum analysis fit is generated [2], which is inspected for randomness in the residuals and to assure a sufficiently low RMSD for each triple. If both criteria are satisfactory, the solutes simulated by the fit represent the original dataset well, and can be used to simulate an equivalent SVE with arbitrary scan times, using the boundary conditions, speeds and hydrodynamic corrections from the original experiment. In order to perform a spectral decomposition, and treat each radial position as a complete wavelength scan, it is important that scans from different wavelengths are time-synchronized. In the Beckman Coulter Optima AUC instrument, each wavelength is acquired sequentially, i.e., at a different time point during the sedimentation process. In order to generate a time-synchronized dataset, each dataset is simulated with its best fit model, shifting all corresponding scans to identical time points. With those simulations in hand, a 3-dimensional surface (absorbance as a function of wavelength, radius and time) is generated for each scan time as shown in SF 4.2. Simulations are accomplished using the ASTFEM simulation module of UltraScan [3, 4], and are based on the best fit model for each triple. At this point, each radial position from each simulated scan represents a wavelength scan that can be deconvoluted into its basis spectra, generating two or more separate two-dimensional datasets (absorbance as a function of radius and time), one for each spectral species. The two-dimensional datasets can then be independently fitted for each species to derive sedimentation and diffusion distributions for each individual analyte, in this case siRNA and LNP.



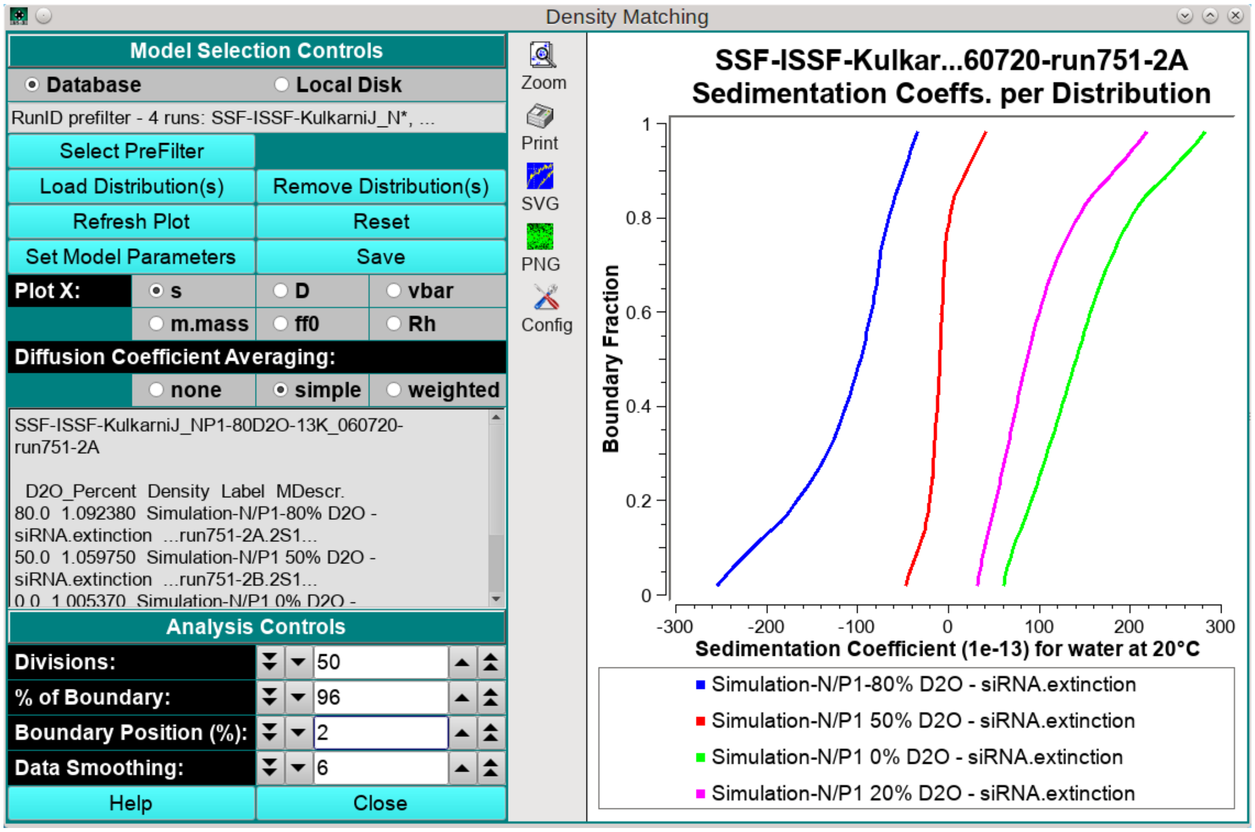
SF 4.2: A single multi-wavelength scan for a LNP experiment before simulation for time synchronization. The time distortion, while small, is resolved by simulating this surface with identical time stamps for each wavelength. The leading peak on the left represents the meniscus. For the time-synchronized surface, each radial position yields a complete wavelength scan that can be decomposed into its spectral basis vectors, generating a two-dimensional dataset for each basis spectrum [16].



Modify Model D2O Percent, Density, Label

Model ndx	D2O Percent	Density (g/l)	Label	Description
Model 1	80	1.00537	Simulation-N/P1-80% D2O - siRNA.extinction	run751-2A.2S1.e2007302227_a2007302228_2DSA-CG_029323_i01
Model 2	50	1.00537	Simulation-N/P1 50% D2O - siRNA.extinction	run751-2B.2S1.e2008121811_a2008121811_2DSA-CG_029348_i01
Model 3	0	1.00537	Simulation-N/P1 0% D2O - siRNA.extinction	run748-2A.2S1.e2007302156_a2007302156_2DSA_029319_i01
Model 4	20	1.00537	Simulation-N/P1 20% D2O - siRNA.extinction	run748-2B.2S1.e2007302212_a2007302213_2DSA_029321_i01

SF 4.3: UltraScan dialog for entering the D_2O percentage for each dataset used in the extrapolation for the PSV distribution.



● Database ○ Local Disk

RunID prefilter - 4 runs: SSF-ISSF-KulkarniJ_N*, ...

Plot X: ● s ○ D ○ vbar ○ m.mass ○ ff0 ○ Rh

Diffusion Coefficient Averaging: ○ none ● simple ○ weighted

```
SSF-ISSF-KulkarniJ_NP1-80D2O-13K_060720-
run751-2A

 D2O_Percent  Density  Label  MDescr.
80.0  1.092380  Simulation-N/P1-80% D2O -
siRNA.extinction  ...run751-2A.2S1...
50.0  1.059750  Simulation-N/P1 50% D2O -
siRNA.extinction  ...run751-2B.2S1...
0.0  1.005370  Simulation-N/P1 0% D2O -
```

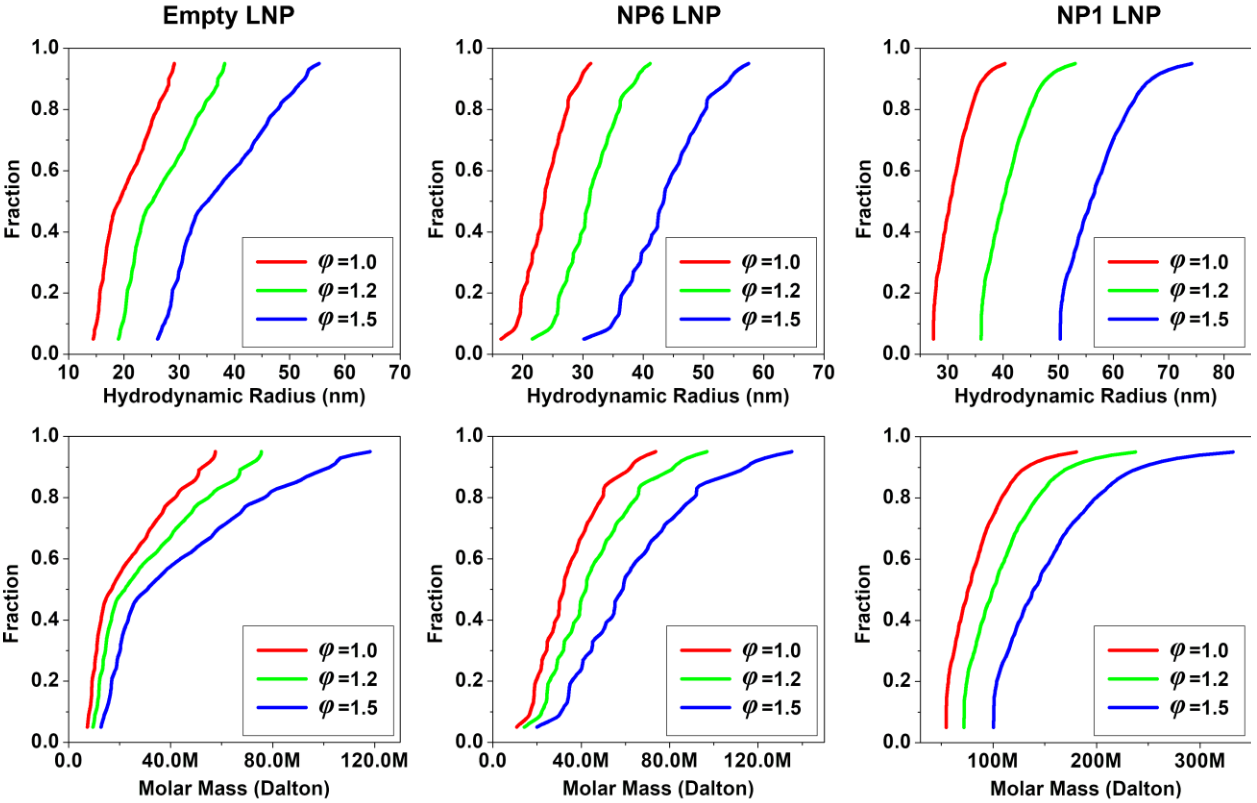
Divisions: 50

% of Boundary: 96

Boundary Position (%): 2

Data Smoothing: 6

SF 4.4: Main screen of the UltraScan us_buoyancy module. Controls for loading distributions and selecting calculated parameters for the integral distributions are shown. Users can switch between displaying s, D, vbar/PSV, molar mass, frictional ratio and hydrodynamic radius. Diffusion coefficient averaging can be performed by using a simple average from all densities, or a weighted average when multiple measurements at the same density are included.



SF 4.5: Hydrodynamic radius (top row) and molar mass (bottom row) predictions for anisotropy constraints 1.0 (red), 1.2 (green) and 1.5 (blue). Comparisons of hydrodynamic radii predicted by DLS suggest that φ = ~1.0 is most consistent, which is also supported by the morphology observed in cryo-TEM images.

Supplemental References

1. Brookes E, Cao W, Demeler B. A two-dimensional spectrum analysis for sedimentation velocity experiments of mixtures with heterogeneity in molecular weight and shape. Eur Biophys J. 2010;39(3):404-414.

2. Demeler B, Brookes E. Monte Carlo analysis of sedimentation experiments. Colloid and Polymer Science. 2008;286:129-137.

3. Cao W, Demeler B. Modeling analytical ultracentrifugation experiments with an adaptive space-time finite element solution of the Lamm equation. Biophys J. 2005;89(3):1589-1602. Epub 2005/06/28. doi: 10.1529/biophysj.105.061135. PubMed PMID: 15980162; PubMed Central PMCID: PMCPMC1366663.

4. Zhang J, Pearson JZ, Gorbet GE, Cölfen H, Germann MW, Brinton MA, et al. Spectral and hydrodynamic analysis of West Nile Virus RNA–protein interactions by multiwavelength sedimentation velocity in the analytical ultracentrifuge. Anal Chem. 2017;89(1):862-870.

Chapter 5. Evaluation of multi-wavelength analytical ultracentrifugation for the characterization and quantification of AAV capsid loading states with UltraScan

Amy Henrickson[1], Xiaozhe Ding[2], Austin G. Seal[3], Zhe Qu[2], Lauren Tomlinson[4], John Forsey[4], Viviana Gradinaru[2], Kazuhiro Oka[3, 5], Borries Demeler[1,6]

[1] Department of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, AB, T1K 3M4, Canada

[2] Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA

[3] Gene Vector Core, Advanced Technology Cores, Baylor College of Medicine Houston, Texas, USA

[4] Pharmaron Biologics Ltd, Speke, Liverpool, UK

[5] Department of Molecular & Cellular Biology, Baylor College of Medicine, Texas, USA

[6] Department of Chemistry and Biochemistry, University of Montana, Missoula, MT, USA

5.1 Contribution of Authors

This chapter consists of a manuscript that was accepted by *Nanomedicine* on September 5, 2023, prepared in collaboration with Dr. Xiaozhe Ding, Austin G. Seal, Dr. Zhe Qu, Dr. John Forsey, Dr. Viviana Gradinaru, and Dr. Kazuhiro Oka who all contributed AAVs. Dr. Xiaozhe Ding also performed TEM at Caltech and Dr. Lauren Tomlinson performed the mass photometry. I performed and analyzed the AUC data, and with the assistance of Dr. Borries Demeler, we conceptualized the experiments and wrote the manuscript.

5.2 Abstract

We present improved methods based on multi-wavelength analytical ultracentrifugation (AUC) to characterize and quantify the presence of empty, partially filled, and filled adeno-associated virus capsids. Currently used AUC methods suffer from low throughput, require relatively large quantities of samples, and have a tendency to overestimate the full capsid quantity. Our methods, multi-wavelength sedimentation velocity, and analytical buoyant density equilibrium address these challenges and significantly improve the accuracy and sensitivity of AUC. They provide high statistical certainty through bulk observations, achieve unambiguous quantification of empty, partially filled, and full capsids, and identify contaminants. We provide guidelines for experimental designs of both AUC methods and compare our results with transmission electron microscopy and mass photometry.

5.3 Introduction

Gene therapy is gaining increased attention for its potential to treat or cure a broad range of genetic diseases. Over the past few years, several gene therapy products have been approved by the European Medicines Agency and the Food and Drug Administration (FDA), while thousands more are candidates in clinical trials [1-3], demonstrating their importance in today's therapeutic arsenal. One of the most established vector systems used is the adeno-associated virus (AAV) [4] due to its minimal pathogenicity and broad clinical applicability [5], resulting in three FDA approved AAV gene therapies to date [2]. The AAV capsid is made up of three proteins VP1, VP2, and VP3, which are typically arranged in a 1:1:10 ratio, with a total of 60 protein subunits [6]. After the capsid is formed, the rAAV genome containing the gene of interest is packaged as ssDNA into the capsids, resulting in full AAVs. Notably, this method relies on the efficient transfection of three or more unique plasmids into a single cell, causing unwanted byproducts of empty and partially filled capsids. Product-related impurities, such as empty capsids, a critical quality attribute, can induce antigen presentation in the cell, leading to increased immune responses and proteasomal degradation of viral capsids, potentially leading to therapeutic failure [7], while also competing with full capsids for cell receptor entry. Therefore, it is essential to accurately characterize AAV loading states, and methodology must be developed to a) efficiently purify full capsids from empty, aggregated, and partially filled capsids, b) quantify gene therapy drug formulations for the precise amount of full capsids compared to the other loading states, and c) identify contaminants. Previously developed analytical methods are limited in their ability to characterize or quantify product-related impurities [8]. Challenges encountered in the characterization of AAV formulations relate to the accuracy and resolution with which this quantification can be made, the throughput and cost of the analysis, and the sensitivity with which the measurement can be made [9]. Other AAV characterization techniques that focus on the characterization of the capsid loading states include transmission electron microscopy (TEM), size exclusion and anion exchange chromatography (SEC/AEC), and mass photometry (MP) [10]. PCR and ELISA are used for the quantification of capsid titers [11]. Each method provides satisfactory results, but with various limitations [10, 12, 13]. The ability of analytical ultracentrifugation (AUC) to characterize and accurately quantify the composition of an AAV sample in a single experiment with minimal sample preparation is a significant advantage of AUC [14]. AUC can detect empty, partially filled, and filled capsids, as well as contaminants, such as free nucleic acids, DNA length heterogeneity in partially filled capsids, aggregates, and full capsids with nucleotide loads fractionally greater than one genome [15]. In addition, the UltraScan software is designed to adhere to 21 CFR Part 11 requirements, allowing AUC to be integrated in a cGMP environment [16].

Current AUC methods primarily employ sedimentation velocity (SV) experiments for AAV characterization [14, 17]. SV measures the transport of macromolecules dissolved in solution, derives the sedimentation and diffusion coefficients of all species in a mixture, and reports their partial concentrations, buoyant molar masses, and shape factors. Despite the popularity of AUC in characterizing empty/full capsid ratios, several shortcomings exist. First, the most often cited SV design only measures between 1 to 3 wavelengths, generally 230, 260, and 280 nm, and at times is matched with Rayleigh interference experiments [17]. These methods overestimate filled capsids due to protein and nucleic acid spectral overlap [18]. Therefore, quantification is always distorted when the DNA and protein extinction coefficients are not separated [19], which is only possible when measuring in multi-wavelength mode [20]. Secondly, SV is a low throughput technique, and analysis is considered difficult, requiring significant training and time investment.

Thirdly, SV experiments require relatively large amounts of sample (~0.5 ml/experiment, 0.3-0.9 OD at 260-280 nm). In an effort to reduce sample requirements, analytical band centrifugation has recently been explored as an alternative to characterize AAV composition [21, 22].

In this study, we report improvements to AUC methods used for the characterization and quantification of AAVs. Specifically, we applied multi-wavelength (MW) analysis to SV and analytical buoyant density equilibrium (ABDE) experiments[1] with AAVs. With the availability of new and improved AUC instruments, MW analysis has been demonstrated to be a powerful new tool for the characterization of the interactions between two or more molecules with distinct absorbance profiles [19, 20, 23-26]. AAV consists of a protein coat and encapsulated DNA, two molecules with unique, but overlapping chromophores. Hence, MW-AUC is ideally well-suited to characterize and quantify the composition of AAV samples. MW-AUC methods take advantage of the optical system in the latest Beckman Optima AUC and new software modules developed for UltraScan [27]. The value of SV analysis performed at multiple wavelengths, and manually analyzed, in a labor-intensive approach was previously demonstrated for the analysis of AAVs by Maruno et al. [22, 28], pointing to the potential of this approach. Recently developed UltraScan software modules now largely automate the MW-AUC analysis workflow, reducing the requirement for training and time spent analyzing AUC data [16, 29-31].

MW-AUC offers a second spectral dimension for the separation of macromolecules in solution in addition to the traditional hydrodynamic separation, and can therefore resolve and identify loading states, distinguish full capsids from contaminants, and resolve unique species with high precision. SV experiments provide separation based on mass, size, anisotropy, and density. SV offers access to a large particle size range and excels at detecting even small

---

[1] Also known as density gradient equilibrium (DGE) experiments

amounts of contaminants. ABDE experiments separate AAV loading states in a CsCl gradient based only on their density. ABDE experiments offer high resolution, sensitivity, and throughput, with minimal sample requirements, and rapid data analysis [18, 29, 32]. Below, we apply multi-wavelength capabilities to SV and ABDE experiments for the analysis of AAV capsid loading states and compare our results to TEM and MP measurements. We show that both MW-AUC methods achieve excellent agreement with TEM and MP, although AUC offers improved statistics due to bulk observation, and enhanced resolution while reducing sample requirements.

5.4 Methods

5.4.1 AAV production

AAV8 was prepared by the Gene Vector Core at Baylor College of Medicine, Houston, TX, using a published method [33]. HEK293T cells were maintained in DMEM (cat# CM002-050, GenDEPOT) supplemented with 10% fetal bovine serum (Seradigm), and antibiotic-antimycotic (cat# CA002-100, GenDepot) at 37°C, 5% CO_2 in a humidified CO_2 incubator. AAV was packaged by a three-plasmid transient transfection with iMFectin Poly DNA Transfection Reagent (cat# I7200, GenDEPOT). A total of 8 µg of DNA was combined with iMFectin and 1 mL of the DNA cocktail was overlayed per dish. Cell associated and media secreted AAVs were collected separately 72 hours after transfection. Cell associated AAVs were recovered by cell lysis using 5% deoxycholate and subsequently treated with DNase I and RNase. Media secreted AAVs were precipitated with 40% PEG8000 supplemented with 2.5M NaCl to a final concentration of 8% PEG. Both AAV fractions were combined and purified by CsCl ultracentrifugation as described by Ayuso et al. [34].

AAV9:CAG-GFP was generated at the CLOVER Center, Caltech, by triple transfection of HEK293T/17 cells (ATCC, cat. CRL-11268) [35] and purified using a precipitation-based method. Briefly, three days after transfection, the producer cells were lifted by adding 10mM EDTA to the media. After being spun down at 2000*g* for 10 minutes, viral particles in the cell pellets were purified with the AAVPro purification kit (Takara bio cat. 6675) according to manufacturer's instructions. AAV9:CAG-mNeonGreen samples were produced in HEK293T/17 cells with the triple transient transfection method and then purified by iodixanol gradient ultracentrifugation as previously described [35]. Virus titers were measured using quantitative polymerase chain reaction (qPCR) targeting the WPRE region of the viral genomes [35].

Empty AAV5 was generated by Pharmaron, UK, using the HEK293 cell line and triple transfection. Following harvest, the feedstock is clarified and subsequently purified. The purification process consists of three chromatography steps: Affinity chromatography, cation exchange chromatography, and anion exchange chromatography.

5.4.2 Electron Microscopy

TEM analysis for AAV8 BCM #1 was performed by the Baylor College of Medicine CryoEM Core. Samples were stained on Quantifoil 2/2 200Cu+ 8nm ThinC grids (Quantifoil Micro Tools GmbH, Jena, Germany) using a 2% uranyl acetate solution (Sigma Aldrich, St. Louis, MO). The images were taken by a 200kV JEOL 2200FS Electron Microscope (JEOL Ltd, Japan) using a Direct Electron DE- 20 camera (Direct Electron, San Diego, CA). The images were then analyzed by FIJI Image J software.

AAV9:CAG-mNeonGreen 1 TEM experiments were performed in the Caltech CryoEM center. Samples were stained on Carbon Type-B 200 Cu grids (Ted Pella, California, US) using 3% uranyl acetate solution (Sigma Aldrich, St. Louis, MO). The images were taken by a Tecnai

T12 120 kV Electron Microscope (FEI, Oregon, US) using a Gatan 2k × 2k CCD (Gatan, California, US).

5.4.3 Mass Photometery:

An AAV5 empty sample was acquired in triplicate on a Refeyn Samux Mass Photometer. A microscope coverslip was cleaned consecutively in MilliQ-Water, isopropanol, and MilliQ-Water prior to mounting a silicone gasket on the coverslip. 10 µL of phosphate buffered saline (PBS) were added to a gasket well and the instrument was focused and locked. 10 µL of AAV5 empty reference sample was added to the PBS solution for a final AAV concentration of $2x10^{11}$ particles/mL. The sample in the well was aspirated using a pipette prior to acquisition of data. Analysis of the data was performed in DiscoverMP, where an AAV mass calibration was applied to convert the measured ratiometric contrast (between scattered and reflected light) to mass. Because the masses for both empty and full particles are known, the empty:full ratios can be determined by this method.

5.4.4 Analytical ultracentrifugation with multi-wavelength SV detection:

MW-AUC SV experiments were performed in a Beckman Coulter Optima AUC (Beckman Coulter, Indianapolis, IN) at the Canadian Center for Hydrodynamics at the University of Lethbridge. Samples were diluted to 0.3-0.4 OD at 260 nm with dPBS (Fisher Scientific, 14190144). Standard Beckman 1.2 cm epon-charcol centerpieces, fitted with quartz windows, were filled with 460 µL of sample. In each experiment, a single cell was measured, and loaded with two samples, utilizing both the sample and reference channel, and measured in intensity mode. AUC cells were loaded into an AN60Ti rotor and equilibrated at 4 °C for 1 hour before the rotor speed was increased to 14.5 krpm and scan collection began. A speed between 14.0-14.5 krpm will maximize the number of scans collected, because the flash lamp rate and rotor speed are synchronized at this speed, reducing the scanning time to ~8 seconds per channel of each cell. 4°C was used to minimize the density variation in response to small, local temperature changes, which nearly eliminates sample convection during the run. Convection can occur in experiments where large analytes are measured at low rotor speeds in dilute buffers when the Optima AUC’s temperature control is activated [36]. The reduced temperature also increases the buffer density and viscosity, resulting in slower sedimentation, permitting additional data to be collected in multi-wavelength mode. Intensity data were collected from both channels of a single cell, from 240-290 nm every second wavelength. All data were analyzed using UltraScan 4.0 [27] as described previously [20, 37]. Briefly, each triple (data from a single cell, channel, and wavelength) was analyzed using the two-dimensional spectrum analysis (2DSA) [38, 39], which removes the time and radially-invariant noise in the intensity data, fits the boundary conditions (meniscus and bottom of the cell), and refines the final model using an iterative approach [37]. Residuals from the iteratively refined 2DSA models were examined for randomness and used to generate a time-synchronized sedimentation profile, for each wavelength, which was used for MW deconvolutions [20, 31].

Diffusion-corrected sedimentation coefficient profiles were generated using the enhanced van Holde –Weischet analysis [40] implemented in UltraScan [27] from the deconvoluted SV data. Dual wavelength AUC results were obtained from the same MW-SV experiments, but only using the iterative 2DSA results from the 260 and 280 nm triples. The buffer density and viscosity corrections were calculated with UltraScan using the partial concentration of each buffer component.

5.4.5 Analytical buoyant density equilibrium (ABDE):

ABDE with AAV samples were performed in CsCl density gradients. ABDE experiments provide high resolution information on AAV loading states, while requiring a significantly reduced amount of AAV sample compared to SV experiments. Like SV experiments, ABDE experiments can also be performed in multi-wavelength mode to add spectral information to separate and quantify protein and DNA signals. ABDE experiments distinguish solutes based on their buoyant densities in a density gradient and produce sharp peaks for each solute that are easily quantifiable with the ABDE peak fitter in UltraScan, which calculates the density of each peak based on peak position by calculating the density gradient profile [29]. The peaks will occur at the positions where the solutes' densities equal the density of the gradient. Together, multi-wavelength and density characterization provides an unambiguous, orthogonal AAV capsid characterization method.

ABDE experiments were performed in a Beckman Coulter Optima AUC, at the University of Lethbridge. Samples were diluted between 0.2-0.25 OD at 260 nm (measured in a 1 cm pathlength), with a final CsCl density of 1.42 g/ml, unless mentioned otherwise, and were prepared in 10 mM Tris (pH 8.0) buffer and measured at 60 krpm. 3 mm centerpieces (epon-charcoal, Beckman Coulter, Indianapolis, IN) were assembled in standard Beckman Coulter cell housings, fitted with sapphire windows and one 3 mm spacer above and below the centerpiece to align the centerpiece fill holes with the fill holes of the cell housing. For accurate quantification, the OD needs to remain within the dynamic range of the detector (0.1-0.9 OD for most wavelengths). 3 mm centerpieces serve two purposes: First, by reducing the pathlength from the typical 12 mm cell used in SV experiments to 3 mm, only ¼ of sample volume is required.

Second, a reduced pathlength significantly decreases the radial refraction error generated by the lens-shaped CsCl gradient, which has a high refractive index.

The density difference generated in the gradient between the meniscus and the bottom of the cell depends on the centrifugal force (F), which is given by $F=m\omega^2r$, where m is mass, ω is the angular velocity, and r is the radius. Therefore, faster rotor speeds and longer columns provide the maximum density range and solute separation. In this configuration, eight samples can be measured simultaneously by utilizing both channels of the cells and all four rotor holes. When measuring with the AN50Ti rotor, sixteen samples can be measured in a single run, but only up to 50 krpm. It should be noted, that as the rotor speed is decreased, the gradient becomes shallower, increasing the CsCl concentration near the meniscus compared to high-speed runs. Therefore, when rotor speeds are decreased the concentration of CsCl required to visualize empty capsids will also need to be decreased, otherwise the empty AAVs peak will not separate from the meniscus.

For ABDE experiments, scans were collected in intensity mode at equilibrium, from 236-288 nm every 2nd wavelength for a total of 27 wavelengths. However, because the sample at equilibrium no longer changes its distribution in the gradient, unlimited time exists to collect up to 100 wavelengths/channel, which is possible with UltraScan's data acquisition software for the Optima AUC. Data collection was performed throughout the experiment to determine when equilibrium was reached, but only the last scan from each wavelength, after the CsCl gradient was established, was used in the analysis. The gradient was deemed established after the solute peaks stopped shifting, approximately after 5 hours at 60 krpm and at 20 °C. Viral particles are relatively large compared to typical proteins, resulting in smaller diffusion coefficients and narrower peaks in an ABDE experiment. This improves baseline resolution between peaks and reduces the amount of sample needed up to 40-fold compared to SV experiments. For ABDE experiments, all data of interest are time invariant. Hence, time- and radially-invariant noise (TI/RI) corrections as performed in SV experiments [41] cannot be used. Therefore, a new noise processing module was developed in UltraScan, called Pseudo-Absorbance, to process the TI and RI noise contributions. It converts intensity data to pseudo-absorbance data using the intensity fingerprint of the photomultiplier tube in the Optima AUC as a function of radius and wavelength. This process extracts the TI background signal from experimental data for each wavelength without the addition of stochastic noise, avoiding the use of a buffer channel for absorbance subtraction. As a result, the signal-to-noise ratio and precision of ABDE data is significantly enhanced. At the same time, the capacity of the instrument is doubled, because the reference channel can be used for another sample [30]. The effect of the noise processing algorithm is illustrated in Figures 5.1 and 5.2. Figure 5.1 shows the samples approaching equilibrium measured in intensity in panel A. Panel B shows the data after conversion to pseudo-absorbance, but without any noise correction, while panel C shows the data after noise correction, which corrects the baseline of the experiment to 0 absorbance. In Figure 5.2, the effect of time-invariant noise contributions resulting from the photomultiplier tube can be clearly seen in the negative baseline observed near 6.2 cm in the red line. Noise processing in UltraScan restores an accurate baseline and recovers reliable peak amplitudes across the entire radial domain. After pseudoabsorbance conversion, the ABDE experiments are edited by selecting the meniscus and cell bottom positions, and spectrally deconvoluted into the spectral contributors, protein and DNA, as explained below.

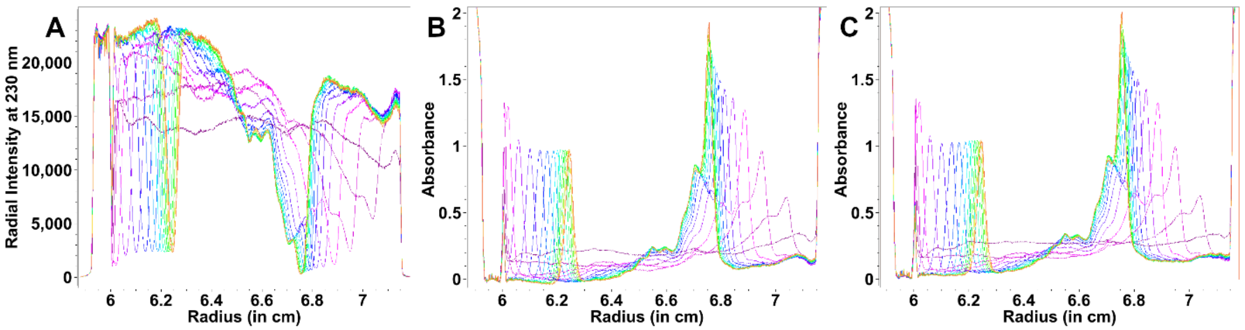


Figure 5.1: Noise correction of an ABDE AAV experiment. A. Raw ABDE intensity data from an AAV sample approaching equilibrium, collected on the Optima AUC (only a single wavelength is shown). B. Pseudo-absorbance conversion of data shown in A without time-invariant noise correction. C. The same data after processing the data with the UltraScan's noise correction module (PseudoAbsorbance). The color gradient indicates time.

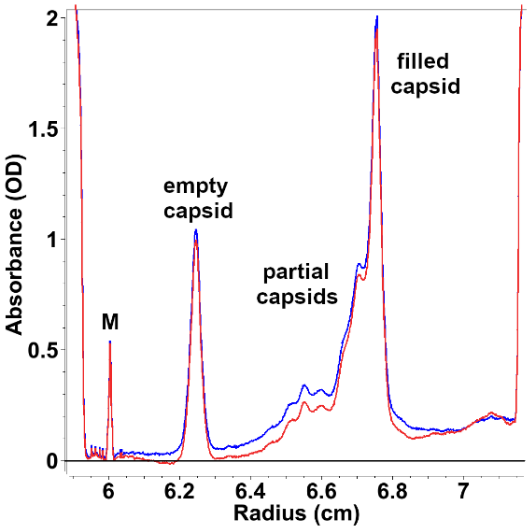


Figure 5.2: Effect of time-invariant noise correction. Shown is the last scan of the ABDE experiment shown in Figure 1. Red: without time-invariant noise correction. Blue: with time invariant noise correction. Note the negative absorbance for the uncorrected scan near 6.2 cm. (M=meniscus position)

The ABDE experiments were analyzed with a recently added module in UltraScan for peak fitting [29]. This program calculates the buoyant density of each analyte from its peak position and other experimental parameters such as the meniscus position, bottom of the cell, gradient forming material properties and concentration, and the rotor speed. Since the density of each analyte is constant, even when any of these experimental parameters are varied, this calculation provides an unambiguous and reproducible characterization of each analyte. Together with the multi-wavelength characterization, this approach provides a very high precision for AAV characterization.

5.4.6 MW Deconvolution

MW-AUC data were deconvoluted into the spectral contributions from each absorbing analyte (protein and DNA) as described in reference [31]. If the buffer or gradient forming material absorbs in the measured wavelength region, it must be included in the spectral deconvolution. The amounts of sample required to measure pure AAV protein and DNA extinction spectra are difficult to obtain. Instead, we chose to measure a dilution series from albumin (Bio Basic cat. D0024, serving as a pure protein standard) and a highly purified DNA plasmid between 240-290 nm. A Genesys 10S benchtop spectrophotometer (Thermo Fisher Scientific) was used to collect all extinction profiles. When measuring between 240-300 nm, the absorbance spectrum from albumin is sufficiently similar to that of the viral capsids because only tryptophan and tyrosine contribute to the spectrum in this range. Hence, they can be used instead of the empty capsid spectrum to obtain a pure spectrum. However, when measuring below 240 nm, the capsid absorbance also depends on the peptide backbone absorption, whose molar absorptivity is primarily a function of the number of peptide bonds in the protein. Because the ratio of the number of peptide bonds and the number of aromatic side chains is not constant, measurements that include wavelengths below 240 nm require a purified empty capsid protein spectrum. A sum of Gaussion terms were fitted to the dilution series and scaled to 1.0 OD at 280 nm for protein and 260 nm for nucleic acid, using the Spectrum Fitter program in UltraScan [27].

Whenever MW-AUC data are deconvoluted into contributions from individual chromophores, the 3D residuals viewer from UltraScan should be used to assess the quality of the decomposition to make sure the basis functions are properly representing the original spectrum [31]. Because accurate molar extinction coefficients are also difficult to obtain for single stranded DNA cargo, packaged inside a protein capsid, we have adopted a different approach for ABDE experiments: We chose to normalize the protein and DNA extinction vectors used for spectral deconvolution such that spectral contributions from protein and DNA observed for the filled capsid species are equal in magnitude. Filled capsids have a known density and are typically the major species in AAV preparations, and hence provide sufficient signal, and are readily identified. Comparison of the normalized absorbance facilitates identification of empty capsids (no spectral contribution from DNA), partially filled capsids (spectral contribution from protein exceeds the DNA contribution) and filled capsids (protein and DNA spectral contributions are equal in amplitude), as well as overloaded species (spectral contribution from DNA exceeds the protein spectral contribution). Precise molar extinction coefficients are generally not available and are not needed to perform the identification of loading states of AAV peaks. Using this approach, it is possible to uniquely determine the relative amount of protein and DNA of each capsid species present in a mixture by comparison to a filled capsid.

5.5 Results

5.5.1 Sedimentation Velocity Experiments:

To determine the composition of AAV formulations, we performed MW-AUC SV experiments, which resolves AAV capsids and contaminants based on their sedimentation coefficient profile and spectral properties, offering excellent resolution. MW-AUC (collected every 2nd wavelength from 240-290 nm) separates the results into the spectral contributors of AAV, the protein capsid and DNA transgene, generating individual hydrodynamic distributions for each macromolecule [34]. The empty-to-full capsid ratio can be directly quantified from the integral sedimentation coefficient distributions generated in the van Holde-Weischet analysis for the MW data [40]. Figure 5.3 compares the 260 nm and 280 nm van Holde-Weischet distributions for an AAV9:CAG-GFP sample with the MW-AUC SV results. Figure 5.3A suggests that 45% of the total concentration sediments as a 100 S species (filled capsids) and 45% as a 65 S species (empty capsids) when measured at 260 nm, while the 280 nm results suggest that 30% of the total concentration sediments as a filled capsid and 65% as an empty capsid. The remainder (10% for the 260 nm measurement, and 5% for the 280 nm measurement) sediment at a lower sedimentation coefficient (< 20 S), and its identity is unknown. The MW-AUC analysis is shown in (Figure 5.3B). Here, the protein and DNA signals are separated into individual van Holde-Weischet distributions, which reflect the relative contribution of each macromolecule in the hydrodynamically separated data. For filled capsids, both DNA and protein signals contribute, while empty capsids only show protein contribution, and partially filled capsids show both protein and DNA signals, with an excess of protein signal. Free nucleic acids are identified by the absence of co-migrating protein signal. Based on the protein signal, the MW-AUC results show that the relative concentration of the 100 S species (filled capsids) contributes only 10% of the total protein signal, while the remaining 90% of the protein signal sediments at 65 S (presumably empty capsids). The DNA signal shows that 65% of the DNA co-migrates with the 100 S species (filled capsids), and 20% co-migrates with the 65 S species. Indicating that the 65 S species is not truly an empty capsid, but contains a small amount of bound nucleic acid, allowing us to classify the 65 S species as a partially filled capsid. The remaining 15% of the DNA signal sediments slower than 20 S, but without any co-migrating protein contribution, identifying the slower species as free nucleic acid. These MW-AUC results demonstrate several key points: 1) Since absorbance spectra for protein and DNA have significant overlaps in the range between 200-300 nm, single wavelength analysis will always add the DNA signal to the signal of the protein species and over-estimate the amount of full capsids. 2) The presence of a small amount of DNA signal in the 65 S species demonstrates that the S-value alone is insufficient to classify capsids as empty, partially or filled capsids, and quantification of AAV capsid loading states using single-wavelength AUC is inherently unreliable. 3) MW-AUC analysis is required to accurately identify any contaminants as proteins, DNA, or protein-DNA complexes. 4) In cases where molar extinction coefficients are available, the protein and DNA distributions will be reported in molar units and MW-AUC can report directly the number and ratio of protein and DNA molecules in each capsid species.

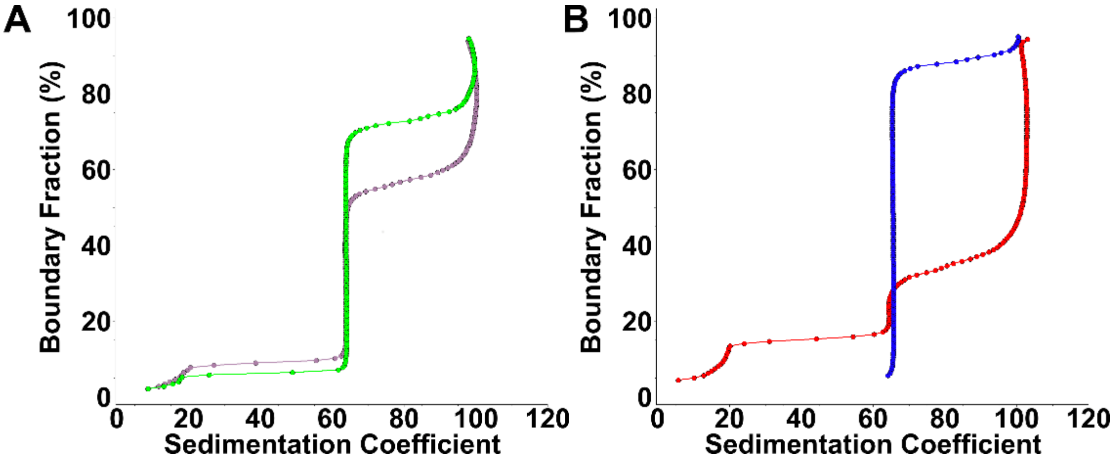


Figure 5.3: Comparison of the enhanced van Holde-Weischet distributions of the 260/280 nm and MW-AUC results: SV AUC data for AAV9:CAG-GFP analyzed using A) only the 260 (Pink) and 280 nm (Green) wavelengths and B) the deconvoluted MW SV AUV method (Blue: 280nm; Red: 260 nm).

5.5.2 ABDE:

ABDE experiments use a density gradient forming material to separate analytes based on their densities, supporting reduced sample volume requirements and increased throughput. The linearity of the ABDE method was evaluated to test the ability to obtain proportional results for different AAV concentrations (see Figure 5.4A and B). For clarity, only the deconvoluted protein signals are shown in Figure 5.4. AAV9:CAG-mNeonGreen 1 was measured at four concentrations ranging between 3.64×10^{12} and 1.36×10^{13} vg/ml (0.11 to 0.40 OD at 260 nm in a

1 cm pathlength). The deconvoluted protein signals from each concentration are shown in Figure 5.4A (without normalization), and their normalized integral distributions are shown in Figure 5.4B, which confirm equal ratios of empty, partially filled, and filled capsids. A minor shift in radial position across all peaks is seen in Figure 5.4A, which results from a slight variation in loading volume. For all concentrations, empty capsid peaks appear at a density between 1.19 to 1.20 g/cm^3 (6.17 to 6.3 cm) and contribute ~15 % to the total signal, and the partially filled capsids contribute ~15 % of the signal and appear between 6.35 and 6.5 cm. The filled capsids appear between 6.5 cm and 6.65 cm and have peak densities between 1.30-1.31 g/cm^3 and contributed ~35 % of the total signal. Unexpectedly, two discrete, higher density species are identified between 6.65 and 6.8 cm with peak densities between 1.37-1.38 g/cm^3 that contribute ~35 % of the signal, which were identified in all samples that contained filled AAVs and is further discussed below.

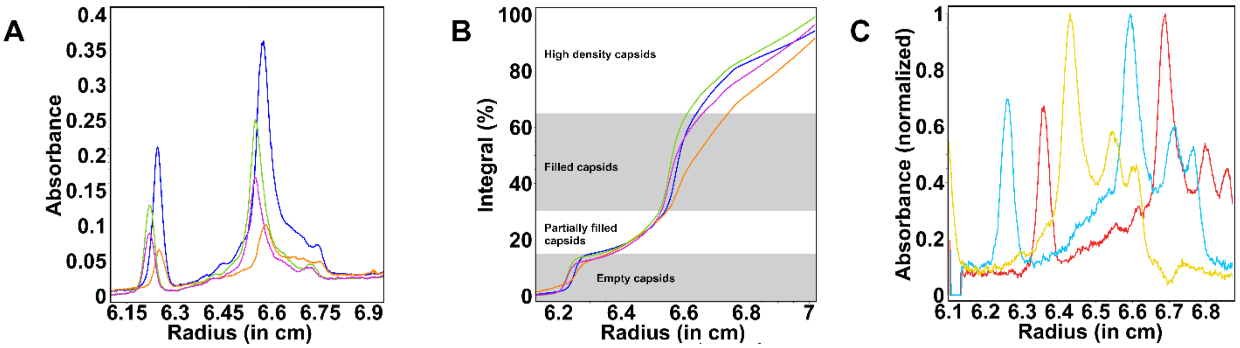


Figure 5.4: ABDE Optimization, only the deconvoluted protein absorption pattern is shown for ABDE experiments for ease of reading. A. Deconvoluted protein signal from ABDE experiments of AAV9:CAG-mNeonGreen 1 at four different AAV concentrations (3.64x1012 vg/ml (black), 6.36x10^{12} vg/ml (green), 9.55x10^{12} vg/ml (pink), and 1.36x10^{13} vg/ml (blue)); B. Normalized integral distribution of the data shown in A, demonstrating that the relative concentrations of each species remain constant. C. normalized results for AAV9:CAG-mNeonGreen 3 in different CsCl concentrations (1.40 g/ml (red), 1.42 g/ml (cyan), and 1.44 g/ml (yellow)).

To optimize the run time and AAV peak formation, various CsCl concentrations, rotor speeds, and CsCl column lengths were tested. The effect of varying CsCl concentrations on peak distributions for sample AAV9:CAG-mNeonGreen 2 is seen in Figure 5.4C. As the CsCl concentrations is increased from 1.35 g/ml (blue), 1.36 g/ml (pink), to 1.38 g/ml (green), all AAV peaks shift to the left. At a CsCl concentration of 1.36 g/ml an optimal distribution of AAV peaks was observed (at 60 krpm), because it positions empty, partially filled, and filled capsids in the center of the solution column, providing baseline separation between the empty and full species, as well as the meniscus and bottom of the cell. At the highest CsCl concentration, the peak corresponding to empty capsids was no longer visible and floated to the top of the cell. At all three CsCl concentrations, the apparent density was ~1.20 g/cm^3 for empty AAVs and between 1.30-1.43 g/cm^3 for the filled AAV capsid and higher density peaks.

Rotor speed can also be used to modulate the CsCl gradient profile. In Figure 5.5A, an overlay of the CsCl equilibrium density gradient profile was simulated for three rotor speeds (40, 50 and 60 krpm), demonstrating the variation in the density profile. Sample AAV9:CAG-mNeonGreen 1 was measured at 50 and 60 krpm to demonstrate the resulting variations in peak positions (Figure 5.5B). At 50 krpm, the gradient is shallower and tends to broaden the peaks, especially towards the bottom of the cell. At 50 krpm, the gradient requires nine hours to reach equilibrium (four hours longer than at 60 krpm). The reduced speed slightly shifts the empty capsids to the left, while the filled AAV peaks are shifted to the right, closer to the bottom of the cell. Although a reduction in rotor speed reduces the density range over which AAV samples can be measured, this approach increases baseline separation and allows the investigator to focus on a particular density range with higher resolution.

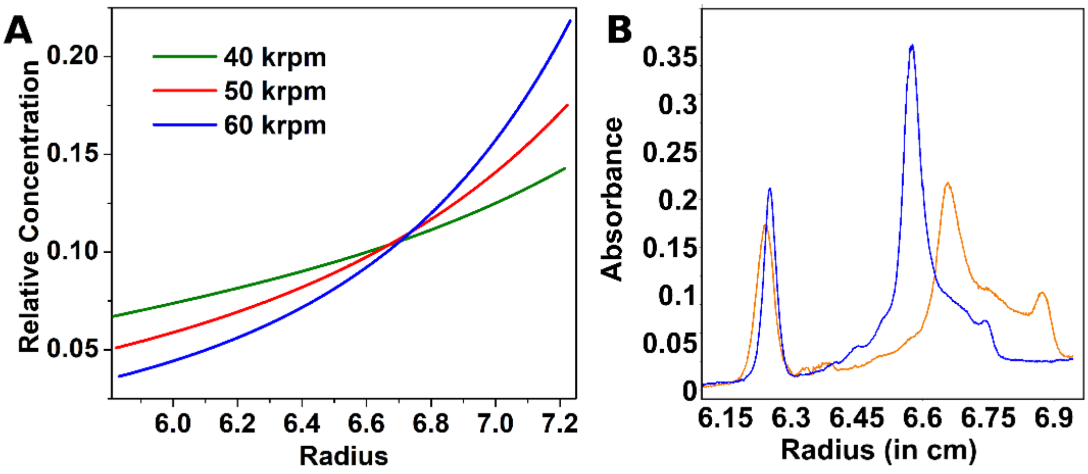


Figure 5.5: Effects of rotor speed on CsCl gradient formation A) Simulated gradient shapes for CsCl at 40 000 rpm (green), 50 000 rpm (red), and 60 000 rpm (blue). B) AAV9:CAG-mNeonGreen 1 measured at 60 000 rpm (blue) and 50 000 rpm (red).

The effect of CsCl column length was simulated with the ASTFEM simulation routine in UltraScan [42, 43], calculating the CsCl distribution for a long (13 mm) column and a short (3 mm) column (Figure 5.6A). The variation in column length clearly demonstrates that at 60 krpm, the CsCl gradient distribution in the long column (13 mm) provides a 3.6 fold higher dynamic density range than the short column (3 mm). Although this approach was advocated by Sternisha et al. [32], we do not recommend to use short column centerpieces for ABDE experiments involving AAV, because the entire range of capsid loading states cannot be distinguished by this approach. The ratio of the maximum density over the minimum density in a 13 mm column is 5.5, but only 1.5 for a 3 mm column. When measuring AAV9:CAG-mNeonGreen 1 in the 3 mm column, at 1.36 g/ml CsCl the sample did not sufficiently separate from the meniscus; therefore, the CsCl density was reduced to 1.35 g/ml. At 1.35 g/ml, in contrast to the long column experiments, only a single peak, likely corresponding to the filled capsid species, was observed in the short column experiments (Figure 5.6B and C). Therefore, the dynamic range of the 3 mm column experiments is insufficient to simultaneously detect empty, partially filled, and filled AAV capsids in a single experiment. Overall, the best experimental design for AAV experiments employed a CsCl concentration of 1.36 g/ml at 60 krpm, 20 ºC, using a 12-13 mm solution column in a 3 mm centerpiece, and measuring wavelengths 240-290 nm. It is recommended to use optically pure CsCl that does not absorb, such as UltraPure™ Cesium Chloride, Optical Grade (Thermo Fisher, 15507023). In addition to background absorbance from CsCl, the high refractive index of CsCl can also contribute to background signal which must be subtracted. To correctly account for the baseline contribution from any absorbance or refractive effects it is advised to also measure the signal from a solution with an identical CsCl concentration in a separate channel. This reference channel absorbance can be used for all other channels in the same run that are performed at the same CsCl concentration. The baseline contribution from CsCl can then be removed from all samples in the experiment using the pseudo-absorbance program in Ultrascan [30].

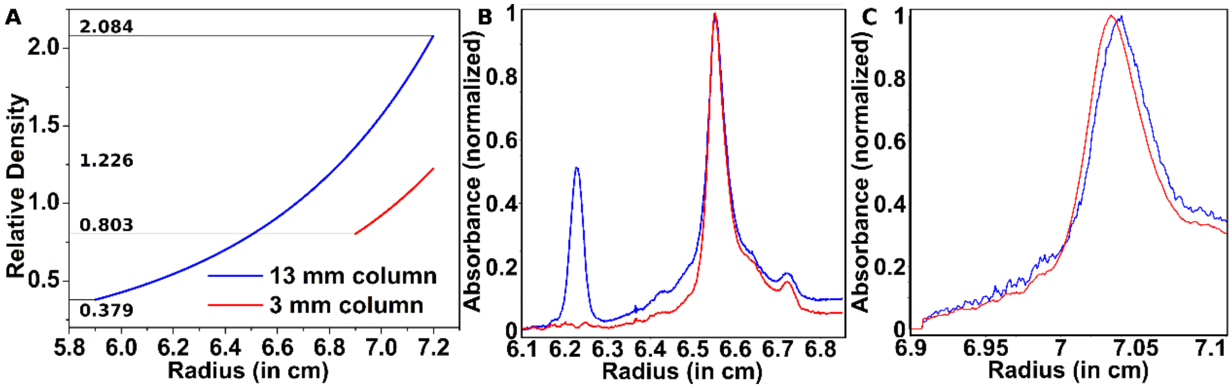


Figure 5.6: Short column vs long column AAV experiments A) Simulated CsCl gradient formation in a 13 mm column (blue) and 3 mm column (red). CsCl gradients from long column experiments (13 mm) provide a 3.6 fold higher dynamic density range than short column experiments (3 mm). B) 13 mm column experiment at 1.42 g/cm^3 CsCl, C) 3 mm column experiment at 1.40 g/cm^3 CsCl

5.5.3 Comparison of multi-wavelength SV, ABDE and TEM:

The ABDE, SV, and TEM results were compared to demonstrate their equivalence. SV analysis of sample AAV8 BCM #1 revealed filled capsids sedimenting at 95 S (Figure 5.7A), along with a nucleic acid signal sedimenting below 20 S. The ABDE results (Figure 5.7B) show a major peak at 6.54 cm with a density of 1.29 g/cm^3, corresponding to the filled capsids. Once again, two smaller higher density peaks are found, at 6.66 cm and 6.74 cm corresponding to densities of 1.34 g/cm^3 and 1.38 g/cm^3, respectively. When the deconvoluted protein and DNA signals are normalized to the amplitude of the filled capsid, it is clear that the two higher density peaks share an identical protein:DNA ratio with the filled capsid peak. The identity of these higher density peaks is not known. One possibility is that they represent overfilled capsids containing more than one complement of the DNA genome. This possibility was rejected because the multi-wavelength analysis of both samples clearly demonstrated identical protein:DNA ratios in the higher density peaks compared to the filled capsid peaks (Figure 5.7A). A larger capsid, accommodating more proteins and DNA is also unlikely because the velocity data of AAV8 BCM #1 does not indicate the presence of any faster sedimenting species. We speculate that the higher density species are the result of a leaky capsid, allowing a small amount of CsCl to enter the capsid, binding to the DNA. The second species could be a capsid containing externalized DNA, or DNA or RNA fragments attached to the exterior of the capsids, freely complexed by CsCl, increasing the overall density of each species. For the TEM analysis, 1460 particles were counted (Figure 5.7C), and only two empty particles were identified.

SV (Figure 5.7D) analysis of AAV9:CAG-mNeonGreen 1 revealed a heterogeneous composition consisting of 55% filled capsids (95 S), 15% empty capsids (~60 S), and ~20% partially filled capsids (70 to 93 S) (Figure 5.7E). ABDE analysis of the same sample revealed a peak containing mostly protein signal, with a density of 1.20 g/cm^3 at 6.25 cm (12.5% of the protein signal), corresponding to empty capsids. Between 6.35 and 6.54 cm, the protein and DNA absorbance gradually increases but never reaches a 1:1 ratio as observed in the filled capsids (6.57 cm). This suggests the presence of a heterogeneous mixture of partially filled capsids containing varying lengths of nucleic acids. The remaining three high density peaks, between 6.55 and 6.77 cm (corresponding to a density of 1.30 g/cm^3 to 1.38 g/cm^3, respectively) all exhibited a 1:1 ratio of protein:DNA. Together these peaks contribute ~51% of the protein signal, similar to the ratio of filled capsids identified by SV. By TEM, the sample was identified to contain ~20% empty capsids and ~80% filled capsids, but the method was unable to distinguish partially filled capsids (Figure 5.7F), likely counting partially filled capsids as filled capsids. Finally, an AAV5 preparation of empty capsid was compared by SV, ABDE, and mass photometry (Figure 5.8). The sedimentation velocity results showed a sedimentation coefficient of ~65 S, the ABDE experiment resulted in a single peak at 6.19 cm and a density of 1.20 g/cm^3. Mass photometry resulted in a single peak with a mass of 3,797 kDa.

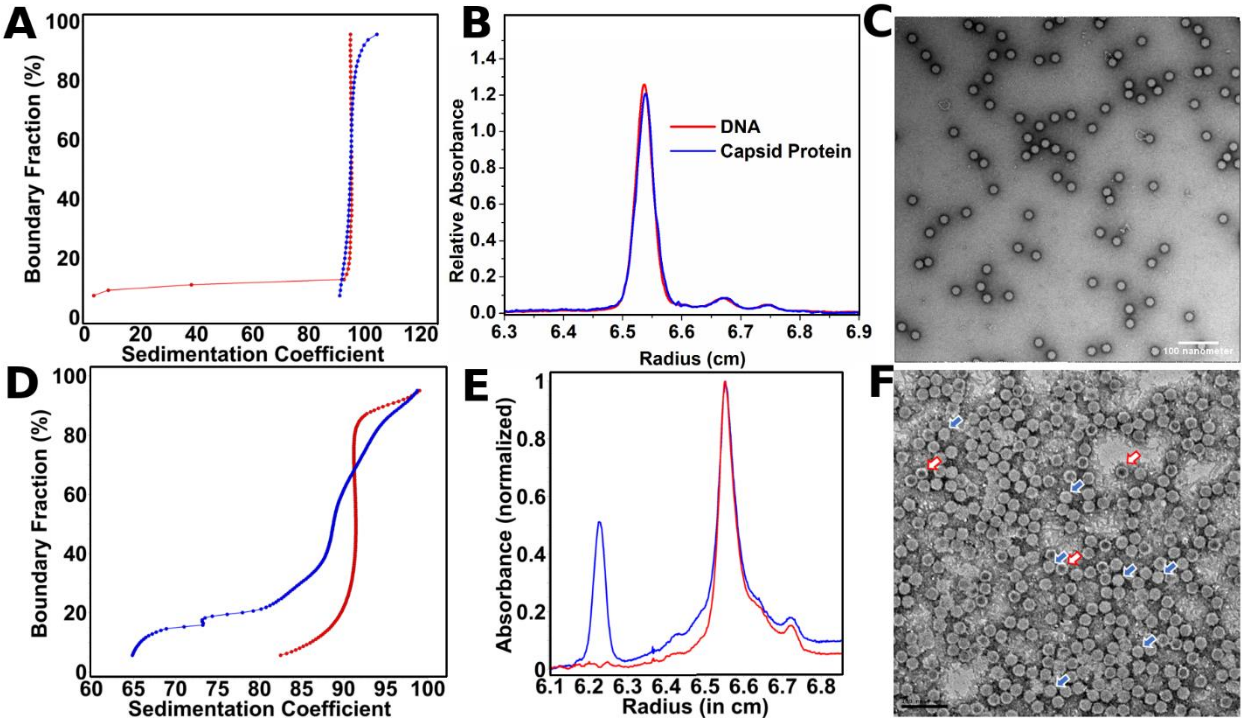


Figure 5.7: AAV formulations measured using SV and ABDE MW-AUC, deconvoluted into its protein absorbance pattern (blue) and DNA absorbance pattern (red), and by TEM. The top panel shows AAV8 analyzed by A) SV MW-AUC, B) ABDE MW-AUC with protein and DNA deconvolutions normalized to 1.3 OD, and C) TEM, no empty AAVs are visible. The bottom panel shows AAV9:CAG-mNeonGreen 2 analyzed by D) SV MW-AUC, E) ABDE MW-AUC with protein and DNA deconvolutions normalized to 1 OD, and F) TEM, the blue arrows show full AAVs and the red arrows empty AAVs.

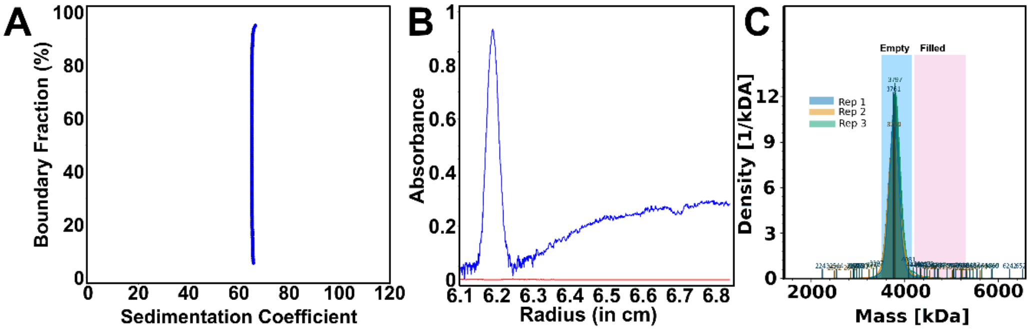


Figure 5.8: Comparison of AAV5 analysis, by A) SV MW-AUC and B) ABDE MW-AUC. For both after the spectral deconvolution only the protein signal (blue) was present. C) shows the results from mass photometry measured in triplicate

5.6 Discussion

Currently, the only well-established method with sufficient resolving power to detect empty, partially full, and full AAV capsids in a single experiment is AUC [44, 45]. However, single wavelength AUC exhibits deficiencies in quantification and identification of samples species. In this study, we describe two improved AUC methods, SV and ABDE, leveraging multi-wavelength detection, which can greatly improve accuracy, resolution, and throughput, and significantly reduce AAV sample requirements. Data acquisition and analysis with UltraScan automates the AUC data acquisition, analysis, and reporting workflow, and supports 21 CFR part 11 requirements to facilitate GMP integration. In addition, high-performance computing possible with UltraScan greatly accelerates data processing. The results from these methods were compared to traditional 260/280 nm AUC experiments, cryo-TEM, and mass photometry measurements, demonstrating significant improvements in accuracy and statistics. Finally, our study demonstrates that MW-AUC methods can be applied to multiple AAV serotypes and transgenes without modification and potentially with minor modifications to rotor speed, run length, and CsCl concentration, to other viral vector systems. Our improvements addressed multiple limitations that prevent the widespread adoption of AUC methods for AAV characterization. The high AAV sample requirement for SV experiments is the most critical limitation. However, it should be noted that sample recovery after SV experiments is possible to permit further testing. By switching to ABDE experiments, the required sample amount is significantly reduced by 20-40 fold. Two factors contribute to the significant sample reduction: The first factor relates to the differences by which the solutes are distributed throughout the cell in each experiment. In any AUC experiment, the measured signal should ideally match the dynamic range of the detector. The sample requirements to reach this limit are very different for the two experiments. In SV experiments, the loading concentration is uniformly distributed over the entire column length. In ABDE experiments, at equilibrium, the entire signal is concentrated into one or more narrow peaks. This concept is illustrated in Figure 5.9, where the orange shaded regions reflect the sample required to reach an optical density (OD) of 0.4 in either experiment. Secondly, the lower sample requirements are a consequence of the enhanced signal-to-noise ratio achieved through the improved noise processing possible in UltraScan [31]. MW-AUC provides additional data points through observations at multiple wavelengths, increasing the accuracy and amplifying the signal-to-noise ratio through additional datapoints by combining all wavelengths in a global fit which separates the protein and DNA signals [20]. Because data are collected at equilibrium, a filled rotor with sixteen samples can be measured at up to 100 wavelengths in the Optima AUC, greatly enhancing sensitivity and throughput at the same time.

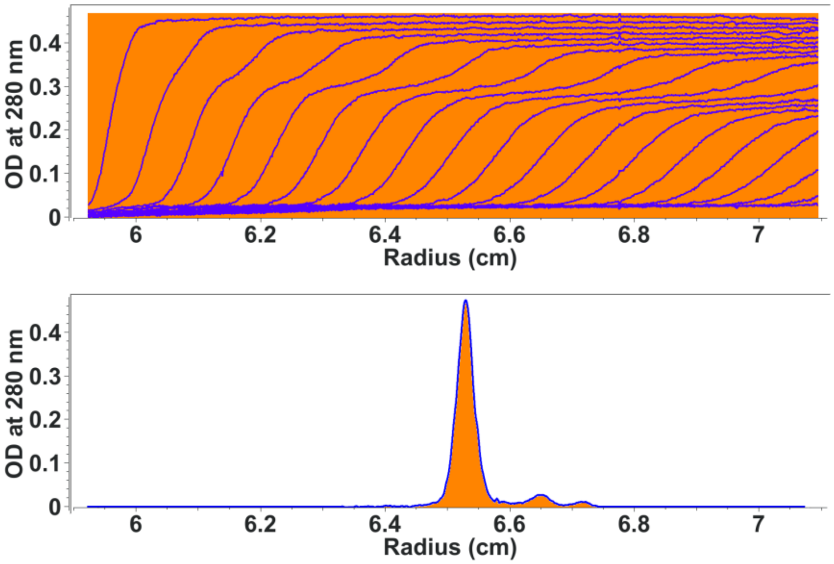


Figure 5.9: Sample quantity requirements – SV vs. ABDE. Required sample amounts are proportional to the orange shaded regions in an SV experiment (top) and an ABDE experiment (bottom). Since ABDE experiments collect the entire signal only in the peak area, sample amounts are significantly reduced. In both experiments, the dynamic range of the detector must be observed to insure a linear response.

MW-AUC offers an orthogonal verification through spectral deconvolution of protein and DNA signals to unambiguously identify and quantify AAV loading states and other contaminants with the highest accuracy and reproducibility. This eliminates uncertainty about filled capsid ratios because the protein capsid can be identified without contribution from the DNA signal, providing an unambiguous mass ratio for each species. It also provides the spectral identity of contaminants, which is essential to assure drug purity and patient safety. The ability to simultaneously characterize the presence of contaminating species and quantifying capsid ratios is an important benefit of the MW-AUC approach. Therefore, SV experiments excel at simultaneously addressing a larger range of solute properties, useful for contaminant identification, but are limited by their throughput. For ABDE experiments the relevant information is accessed after equilibrium has been reached, allowing not only more time for data collection at multiple wavelength, but also for collection of additional cells and channels, significantly improving throughput. Our studies included a comparison with negative stain TEM experiments, which showed excellent agreement when only full or empty capsids were present, but the presence of partially filled capsids could not be established by TEM.

Our study demonstrates the relevance and reliability of SV and ABDE MW-AUC for the characterization and quantification of AAV loading states in AAV serotypes 5, 8, and 9. UltraScan also includes 21 CFR Part 11 compliance design requirements, allowing for easier GMP environment integration and sample analysis. Taken together, improved MW-AUC should help reinforce AUC as the gold standard analysis method. ABDE experiments revealed a previously unknown feature of AAV capsids, the presence of two additional higher density peaks with the same protein:DNA ratio as the full capsid (see Figures 5.4 A-C, 5.6 B, and 5.7 B and E). Capsid instabilities under certain conditions have been reported earlier [46, 47], pointing to a possible undesirable species in therapeutic formulations. Further research is warranted to better understand the identity of these two higher density species, and the role of pH and CsCl in the solvent as factors potentially affecting capsid stability.

5.7 References

1. Ma C-C, Wang Z-L, Xu T, He Z-Y, Wei Y-Q. The approved gene therapy drugs worldwide: from 1998 to 2019. Biotechnology advances. 2020;40:107502.

2. FDA. Approved Cellular and Gene Therapy Products [updated May 2023]. Available from: https://www.fda.gov/vaccines-blood-biologics/cellular-gene-therapy-products/approved-cellular-and-gene-therapy-products.

3. b HNo. Gene Therapy Clinical Trials. Available from: https://clinicaltrials.gov/ct2/results?cond=&term=gene+therapy&cntry=&state=&city=&dist=.

4. Wang D, Tai PW, Gao G. Adeno-associated virus vector as a platform for gene therapy delivery. Nature reviews Drug discovery. 2019;18(5):358-378.

5. Li C, Samulski RJ. Engineering adeno-associated virus vectors for gene therapy. Nature Reviews Genetics. 2020;21(4):255-272.

6. Wörner TP, Bennett A, Habka S, Snijder J, Friese O, Powers T, et al. Adeno-associated virus capsid assembly is divergent and stochastic. Nat Commun. 2021;12(1):1642. Epub 2021/03/14. doi: 10.1038/s41467-021-21935-5. PubMed PMID: 33712599; PubMed Central PMCID: PMCPMC7955066 in employing AAV vectors for gene delivery purposes. M.A.-M. is co-founder of StrideBio, Inc., with an interest in developing AAV technology for gene delivery purposes. M.A.-M. is a member of the ATGC and Voyager SAB, and consultants for StrideBio, Intima Bioscience, being biopharma companies with interest in developing AAV for gene delivery purposes. The remaining authors declare no competing interests.

7. Pei X, Earley LF, He Y, Chen X, Hall NE, Samulski RJ, et al. Efficient capsid antigen presentation from adeno-associated virus empty virions in vivo. Frontiers in immunology. 2018;9:844.

8. Wright JF. Product-related impurities in clinical-grade recombinant AAV vectors: characterization and risk assessment. Biomedicines. 2014;2(1):80-97.

9. Colomb-Delsuc M, Raim R, Fiedler C, Reuberger S, Lengler J, Nordström R, et al. Assessment of the percentage of full recombinant adeno-associated virus particles in a gene therapy drug using CryoTEM. PLoS ONE. 2022;17(6):e0269139.

10. Wu D, Hwang P, Li T, Piszczek G. Rapid characterization of adeno-associated virus (AAV) gene therapy vectors by mass photometry. Gene Ther. 2022;29(12):691-697. Epub 2022/01/21. doi: 10.1038/s41434-021-00311-4. PubMed PMID: 35046529; PubMed Central PMCID: PMCPMC9296698.

11. Gimpel AL, Katsikis G, Sha S, Maloney AJ, Hong MS, Nguyen TNT, et al. Analytical methods for process and product characterization of recombinant adeno-associated virus-based gene therapies. Mol Ther Methods Clin Dev. 2021;20:740-754. Epub 2021/03/20. doi: 10.1016/j.omtm.2021.02.010. PubMed PMID: 33738328; PubMed Central PMCID: PMCPMC7940698.

12. Mayginnes JP, Reed SE, Berg HG, Staley EM, Pintel DJ, Tullis GE. Quantitation of encapsidated recombinant adeno-associated virus DNA in crude cell lysates and tissue culture medium by quantitative, real-time PCR. J Virol Methods. 2006;137(2):193-204. doi: https://doi.org/10.1016/j.jviromet.2006.06.011. PubMed PMID: MAYGINNES2006193.

13. Fagone P, Wright JF, Nathwani AC, Nienhuis AW, Davidoff AM, Gray JT. Systemic Errors in Quantitative Polymerase Chain Reaction Titration of Self-Complementary Adeno-Associated Viral Vectors and Improved Alternative Methods. Human Gene Therapy Methods. 2012;23(1):1-7. doi: 10.1089/hgtb.2011.104. PubMed PMID: 22428975.

14. Yarawsky AE, Zai-Rose V, Cunningham HM, Burgner JW, 2nd, DeLion MT, Paul LN. AAV analysis by sedimentation velocity analytical ultracentrifugation: beyond empty and full capsids. Eur Biophys J. 2023. Epub 2023/04/11. doi: 10.1007/s00249-023-01646-z. PubMed PMID: 37037926.

15. Hayes DB, Dobnik D. Commentary: Multiplex dPCR and SV-AUC are Promising Assays to Robustly Monitor the Critical Quality Attribute of AAV Drug Product Integrity. J Pharm Sci. 2022.

16. Savelyev A, Gorbet GE, Henrickson A, Demeler B. Moving analytical ultracentrifugation software to a good manufacturing practices (GMP) environment. PLoS Comput Biol. 2020;16(6):e1007942. doi: 10.1371/journal.pcbi.1007942.

17. Werle AK, Powers TW, Zobel JF, Wappelhorst CN, Jarrold MF, Lyktey NA, et al. Comparison of analytical techniques to quantitate the capsid content of adeno-associated viral vectors. Molecular Therapy-Methods & Clinical Development. 2021;23:254-262.

18. Berkowitz SA, Philo JS. Monitoring the homogeneity of adenovirus preparations (a gene therapy delivery system) using analytical ultracentrifugation. Anal Biochem. 2007;362(1):16-37.

19. Gorbet GE, Pearson JZ, Demeler AK, Cölfen H, Demeler B. Next-generation AUC: analysis of multiwavelength analytical ultracentrifugation data. Methods Enzymol. 562: Elsevier; 2015. p. 27-47.

20. Henrickson A, Gorbet GE, Savelyev A, Kim M, Hargreaves J, Schultz SK, et al. Multi-wavelength analytical ultracentrifugation of biopolymer mixtures and interactions. Anal Biochem. 2022:114728.

21. Khasa H, Kilby G, Chen X, Wang C. Analytical band centrifugation for the separation and quantification of empty and full AAV particles. Mol Ther Methods Clin Dev. 2021;21:585-591. Epub 2021/06/08. doi: 10.1016/j.omtm.2021.04.008. PubMed PMID: 34095342; PubMed Central PMCID: PMCPMC8142049.

22. Maruno T, Ishii K, Torisu T, Uchiyama S. Size Distribution Analysis of the Adeno-Associated Virus Vector by the c(s) Analysis of Band Sedimentation Analytical Ultracentrifugation with Multiwavelength Detection. J Pharm Sci. 2023;112(4):937-946. Epub 2022/11/15. doi: 10.1016/j.xphs.2022.10.023. PubMed PMID: 36374763.

23. Horne CR, Henrickson A, Demeler B, Dobson RC. Multi-wavelength analytical ultracentrifugation as a tool to characterise protein–DNA interactions in solution. Eur Biophys J. 2020;49(8):819-827.

24. Horne CR, Venugopal H, Panjikar S, Wood DM, Henrickson A, Brookes E, et al. Mechanism of NanR gene repression and allosteric induction of bacterial sialic acid metabolism. Nat Commun. 2021;12(1):1988. Epub 2021/04/02. doi: 10.1038/s41467-021-22253-6. PubMed PMID: 33790291; PubMed Central PMCID: PMCPMC8012715.

25. Gabir H, Gupta M, Meier M, Heide F, Koch M, Stetefeld J, et al. Investigation of dynamic solution interactions between NET-1 and UNC-5B by multi-wavelength analytical ultracentrifugation. Eur Biophys J. 2023. Epub 2023/03/21. doi: 10.1007/s00249-023-01644-1. PubMed PMID: 36939874.

26. Ahmed I, Hahn J, Henrickson A, Khaja FT, Demeler B, Dubnau D, et al. Structure-function studies reveal ComEA contains an oligomerization domain essential for transformation in gram-positive bacteria. Nat Commun. 2022;13(1):7724. Epub 2022/12/14. doi: 10.1038/s41467-022-35129-0. PubMed PMID: 36513643; PubMed Central PMCID: PMCPMC9747964.

27. Demeler B, Gorbet GE. Analytical ultracentrifugation data analysis with UltraScan-III. In Analytical Ultracentrifugation: Instrumentation, Software, and Applications; Uchiyama, S., Stafford, W. F., Laue, T., Eds.; . Analytical Ultracentrifugation: Springer; 2016. p. 119-143.

28. Maruno T, Usami K, Ishii K, Torisu T, Uchiyama S. Comprehensive size distribution and composition analysis of adeno-associated virus vector by multiwavelength sedimentation velocity analytical ultracentrifugation. J Pharm Sci. 2021;110(10):3375-3384.

29. Savelyev A, Brookes EH, Henrickson A, Demeler B. A new UltraScan module for the characterization and quantification of analytical buoyant density equilibrium experiments to determine AAV capsid loading. Eur Biophys J. 2023. Epub 2023/04/05. doi: 10.1007/s00249-023-01641-4. PubMed PMID: 37014454.

30. Mortezazadeh S, Demeler B. Systematic noise removal from analytical ultracentrifugation data with UltraScan. Eur Biophys J. 2023. Epub 2023/02/15. doi: 10.1007/s00249-023-01631-6. PubMed PMID: 36786920.

31. Mortezazadeh S, Demeler B. A spectral decomposition quality assessment tool for multi-wavelength AUC experiments with UltraScan. Eur Biophys J. 2023. Epub 2023/03/18. doi: 10.1007/s00249-023-01640-5. PubMed PMID: 36930298.

32. Sternisha SM, Wilson AD, Bouda E, Bhattacharya A, VerHeul R. Optimizing high-throughput viral vector characterization with density gradient equilibrium analytical ultracentrifugation. Eur Biophys J. 2023. Epub 2023/05/03. doi: 10.1007/s00249-023-01654-z. PubMed PMID: 37130969.

33. Deng S, Oka K. Adeno-Associated Virus as Gene Delivery Vehicle into the Retina. Methods Mol Biol. 2020;2092:77-90. Epub 2019/12/02. doi: 10.1007/978-1-0716-0175-4_7. PubMed PMID: 31786783.

34. Ayuso E, Mingozzi F, Montane J, Leon X, Anguela XM, Haurigot V, et al. High AAV vector purity results in serotype- and tissue-independent enhancement of transduction efficiency. Gene Ther. 2010;17(4):503-510. Epub 2009/12/04. doi: 10.1038/gt.2009.157. PubMed PMID: 19956269.

35. Challis RC, Kumar SR, Chan KY, Challis C, Beadle K, Jang MJ, et al. Systemic AAV vectors for widespread and targeted gene delivery in rodents. Nat Protoc. 2019;14(2):379-414.

36. Berkowitz SA, Laue T. Boundary convection during sedimentation velocity in the Optima analytical ultracentrifuge. Anal Biochem. 2021;631:114306.

37. Demeler B. Methods for the design and analysis of sedimentation velocity and sedimentation equilibrium experiments with proteins. Current protocols in protein science. 2010;60(1):7.13. 11-17.13. 24.

38. Brookes EH, Boppana RV, Demeler B, editors. Computing large sparse multivariate optimization problems with an application in biophysics. SC'06: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing; 2006: IEEE.

39. Brookes E, Cao W, Demeler B. A two-dimensional spectrum analysis for sedimentation velocity experiments of mixtures with heterogeneity in molecular weight and shape. Eur Biophys J. 2010;39(3):404-414.

40. Demeler B, Van Holde KE. Sedimentation velocity analysis of highly heterogeneous systems. Anal Biochem. 2004;335(2):279-288.

41. Schuck P, Demeler B. Direct Sedimentation Analysis of Interference Optical Data in Analytical Ultracentrifugation. Biophys J. 1999;76(4):2288-2296. doi: https://doi.org/10.1016/S0006-3495(99)77384-4.

42. Cao W, Demeler B. Modeling analytical ultracentrifugation experiments with an adaptive space-time finite element solution of the Lamm equation. Biophys J. 2005;89(3):1589-1602. Epub 2005/06/28. doi: 10.1529/biophysj.105.061135. PubMed PMID: 15980162; PubMed Central PMCID: PMCPMC1366663.

43. Cao W, Demeler B. Modeling analytical ultracentrifugation experiments with an adaptive space-time finite element solution for multicomponent reacting systems. Biophys J. 2008;95(1):54-65. Epub 2008/04/09. doi: 10.1529/biophysj.107.123950. PubMed PMID: 18390609; PubMed Central PMCID: PMCPMC2426643.

44. Burnham B, Nass S, Kong E, Mattingly M, Woodcock D, Song A, et al. Analytical ultracentrifugation as an approach to characterize recombinant adeno-associated viral vectors. Human gene therapy methods. 2015;26(6):228-242.

45. Fu X, Chen WC, Argento C, Clarner P, Bhatt V, Dickerson R, et al. Analytical Strategies for Quantification of Adeno-Associated Virus Empty Capsids to Support Process Development. Hum Gene Ther Methods. 2019;30(4):144-152. Epub 2019/08/02. doi: 10.1089/hgtb.2019.088. PubMed PMID: 31368356.

46. Bernaud J, Rossi A, Fis A, Gardette L, Aillot L, Büning H, et al. Characterization of AAV vector particle stability at the single-capsid level. Journal of biological physics. 2018;44(2):181-194.

47. Ros C, Baltzer C, Mani B, Kempf C. Parvovirus uncoating in vitro reveals a mechanism of DNA release without capsid disassembly and striking differences in encapsidated DNA stability. Virology. 2006;345(1):137-147.

Chapter 6. Conclusions

Throughout this thesis, I have discussed analytical ultracentrifugation (AUC), a technique that uses light to measure the fundamental relationship between centrifugal force and a macromolecule's size, shape, and density, with a specific focus on MW-AUC. MW-AUC provides a second orthogonal characterization method to resolve analytes based on their optical properties along with their hydrodynamic properties. Chapters 2 and 3 focus on experimental considerations, design, analysis, and utilization of MW-AUC. The basics of MW-AUC, established in these first chapters, are then applied to several other AUC methods in Chapters 4 and 5 to accurately quantify, characterize, and analyze LNP and AAV-based gene therapies.

Chapter 2 focuses on essential considerations regarding designing and analyzing MW-AUC experiments. It demonstrates the ability of MW-AUC to provide useful information on complex samples even if the pure basis spectra of each analyte are unknown. When the pure basis spectra are known, the molar quantities of each analyte can be resolved, as can the stoichiometry and thermodynamic coefficients of the interactions. Chapter 3 uses AUC to elucidate the structure-function relationship of ComEA and determine the K_d and k_{off} of oligomerization. It then applies MW-AUC study the binding of ComEA to various DNA strands, demonstrating the abilities of MW-AUC to characterize systems with reversible interactions. These two chapters are essential to establishing an understanding of MW-AUC.

This understanding of MW-AUC allows the user to apply the MW technique to other AUC methods, such as density matching and ABDE. Using these methods in conjunction with MW optics results in the accurate characterization of gene therapeutics, providing the relative ratio of all analytes and contaminants in the sample while also gaining information on their hydrodynamic properties, such as partial specific volumes and molar masses. Chapter 4 uses MW-AUC and density matching AUC to determine the loading state of LNPs, a non-viral vector. The MW-AUC data highlighted the size heterogeneity of the preparations but also showed that all LNPs contained RNA. The density matching studies were used to experimentally determine the samples' partial specific volume distribution and loading states. After finding the partial specific volume, the molar mass distributions can be calculated. The molar masses and hydrodynamic radii are important qualities to be aware of as they may contribute to the therapeutics' effectiveness [1]. Chapter 5 focuses on the viral vector AAV; it highlights the advantages of MW ABDE experiments and compares the results to SV MW-AUC, dual-wavelength AUC, TEM, and mass photometry. In both SV and ABDE experiments, the use of MW optics results in the accurate determination of loaded, empty, partially filled, and overfilled AAV concentrations and the identification of other contaminants. In addition, ABDE experiments significantly reduce the sample volume required while increasing the throughput and maintaining all the benefits of MW-AUC. When used in conjunction with UltraScan, the apparent partial specific volume or density of each peak can be calculated, increasing the hydrodynamic characterization potential of AUC. Overall, this thesis presented MW-AUC methods that can increase the accuracy of characterization methods for gene therapeutics.

6.1 Future directions

Recently a new UltraScan program, UltraScan-GMP, was developed to help align AUC with the good manufacturing processes (GMP) outlined by the FDA, and a new tool was also added to UltraScan that checks the validity of the optical deconvolutions, given the provided basis spectra [2]. Continuous improvements to the methods and analysis help to ensure that AUC will stay at the top of its field and be able to transition into more research areas. Currently, the drawback of

SV MW experiments with the Beckman optics is that only two samples can usually be measured at a time. A program could be developed to improve the throughput using Equation 2.1 and the pure spectral contributors from the sample to determine the minimal number of wavelengths that can be measured without significantly reducing angle θ between the spectral pairs. If the number of measured wavelengths could be reduced without sacrificing the optical deconvolution quality, it could potentially allow for a second AUC cell to be measured, doubling the capacity of MW-SV experiments.

Further studies should also be conducted on the high density AAV peaks identified in the ABDE experiments, noted in Chapter 5, which are significant as CsCl gradients are used to purify some AAV preparations. These high density species are seen in almost all AAV samples we have analyzed, although they are more prominent in some samples. They have the same ratio of protein to DNA as the filled capsid, meaning that the shift in density is not a result of different loading states. These additional AAV species are also not seen in experiments that use physiological conditions, such as PBS or TRIS, and therefore are likely a result of the CsCl. Thus far, we have some reason to believe that the appearance of these peaks may be time-sensitive and that the longer the AAVs are exposed to CsCl, the more prominent they become. Therefore, a time-course study of the AAVs in the CsCl would be a logical first experiment, helping to determine when the peaks initially occur and if there is an endpoint in the transition. This experiment would provide some useful knowledge, but further experiments must be performed to determine why or how these peaks form. Our current hypothesis is that one peak may result from "leaky" AAVs, which allow CsCl to enter the capsid and complex with the DNA, increasing their density. The second peak may result from the AAVs becoming destabilized in the CsCl and externalizing their DNA. This externalized DNA increases the surface area CsCl can bind to, further increasing their density. The initial tests should determine if the AAV capsids are “leaky” under physiological conditions. To do so a fluorescent dye could be added to AAVs in PBS; if they are leaky, the dye will enter the capsid and bind with the DNA, resulting in fluorescence. Even if no fluorescence is detected in these experiments, a similar experiment should be repeated as an ABDE experiment to determine if the CsCl destabilizes the capsids, which may be possible as the extra AAV species are not seen unless CsCl is present. In the ABDE experiments, if the high density peaks result from the “leaky” or destabilized capsids, only those peaks should fluoresce. If we can determine the cause of the high-density peaks, researchers could look for ways to reduce them and increase their AAV yield.

The AAVs isolated from the high-density peaks should also be studied under physiological conditions to determine if and how they differ from AAVs collected from the filled fraction. AAVs collected from the high density fractions in preparative CsCl gradients should be buffer exchanged into PBS and measured in MW-SV and ABDE experiments. Determining the hydrodynamic properties of the high density AAVs will provide insight into whether these AAVs are hydrodynamically different from those in the filled fraction. If identical, researchers could potentially collect them with the filled AAV fraction, increasing the concentration of AAVs collected. However, more extensive studies would be needed to determine if AAVs in the high density fraction are safe for patient use or if this fraction be avoided at all costs.

6.2 References

1. Belliveau NM, Huft J, Lin PJ, Chen S, Leung AK, Leaver TJ, et al. Microfluidic Synthesis of Highly Potent Limit-size Lipid Nanoparticles for In Vivo Delivery of siRNA. Mol Ther Nucleic Acids. 2012;1(8):e37. Epub 2013/01/25. doi: 10.1038/mtna.2012.28. PubMed PMID: 23344179; PubMed Central PMCID: PMCPMC3442367.

2. Mortezazadeh S, Demeler B. A spectral decomposition quality assessment tool for multi-wavelength AUC experiments with UltraScan. Eur Biophys J. 2023. Epub 2023/03/18. doi: 10.1007/s00249-023-01640-5. PubMed PMID: 36930298.