Characterization of Nucleic Acid Delivery Vehicle Products using Microfluidic Concepts

By

Adriana Coll De Peña

M.S. Science, Technology and Public Policy, Rochester Institute of Technology, 2020

B.S. Biomedical Engineering, Rochester Institute of Technology, 2020

B.S. Molecular Bioscience and Biotechnology, Rochester Institute of Technology, 2017

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Biomedical Engineering at Brown University

PROVIDENCE, RHODE ISLAND

February 2024

This dissertation by Adriana Coll De Peña is accepted in its present form by the Biomedical Engineering program, a joint program in the Division of Biology and Medicine and the School of Engineering as satisfying the dissertation requirement for the degree of Doctor of Philosophy.

Date:_____ _____

Anubhav Tripathi, Ph.D., Advisor

Recommended to the Graduate Council

Date:_____ _____

Eric Darling, Ph.D., Reader

Date:_____ _____

Anita Shukla, Ph.D., Reader

Date:_____ _____

Robert Hurt, Ph.D., Reader

Date:_____ _____

Ian Wong, Ph.D., Reader

Date:_____ _____

Somdatta Goswami, Ph.D., Reader

Date:_____ _____

Christina Bailey-Hytholt, Ph.D., Reader

Approved by the Graduate Council

Date:_____ _____

Thomas A. Lewis, Ph.D., Dean of the Graduate School

Abstract

In recent years, there has been an exponential growth in biopharmaceutical development stemming from its ability to treat previously incurable diseases and to go from pathogen discovery to vaccine creation at unprecedented speeds, as highlighted by the COVID-19 response. Broadly defined as biological pharmaceuticals manufactured using biotechnology, the manufacturing of biopharmaceuticals requires multiple steps of purification and validation that significantly increase the production cost. The validation of the production and different purification steps requires the integration of analytical monitoring throughout the process. However, despite being developed and suitable for the less complex chemical pharmaceuticals, analytical tools have not been capable of evolving at the same rate as the biopharmaceutical field. This mismatch in evolution has left a significant unmet need for rapid, high-throughput analytical methods capable of analyzing the numerous samples generated during the iterative manufacturing of biopharmaceuticals. With its rapid turnaround times, high-throughput compatibility, and low sample requirements, microfluidic electrophoresis presents a promising solution to address this need. Overall, the work described in this thesis increases the understanding of the electrophoretic transport of biopharmaceutical products to inform the development of streamlined microfluidic analytical methods for the purity assessment of samples. To achieve this, we combine a fundamental understanding of the biochemistry and biophysical properties of each sample and exploit these properties to achieve differentiation. The impact of this work includes increasing the electrophoretic understanding of biopharmaceuticals to decrease the analytical burden of their characterization by reducing personnel-intensive steps, turnaround time, reagent and sample volume requirements, and cost through the use of microfluidics and electrophoresis. The application of this approach was shown

with a wide range of nucleic acid delivery vehicle products ranging from nucleic acids to lipids and viruses. Ultimately, we developed translatable technology applicable to numerous biopharmaceutical platforms by investigating the fundamental electrophoretic properties of a wide range of molecules.

Curriculum Vitae

Adriana Coll De Peña

Education

Ph.D. in Biomedical Engineering September 2020 - December 2023
Brown University, Providence, RI

M.S. in Science, Technology and Public Policy January 2018 - August 2020
Rochester Institute of Technology, Rochester, NY
Graduate Commencement Delegate Speaker, College of Liberal Arts

B.S. in Biomedical Engineering August 2017 - May 2020
Rochester Institute of Technology, Rochester, NY

B.S. in Molecular Bioscience and Biotechnology August 2013-May 2017
Rochester Institute of Technology, Rochester, NY
Minor: German

Publications

17. Pollock, J.; **Coll De Peña, A.**; Tripathi, A. Rapid microfluidic aneuploidy screening method. In preparation.

16. **Coll De Peña, A.**; Chen, N.; Vaduva, M.; Bailey-Hytholt, C.M.; Tripathi, A. Application of microfluidic analytical techniques to support the development of biological nucleic acid delivery vehicles. In Preparation.

15. **Coll De Peña, A.**; Vaduva, M.; Li, N.S.; Shah, S.; Ben Frej, M.; Tripathi, A. Enzymatic Isolation and Microfluidic Electrophoresis Analysis of Residual dsRNA Impurities in mRNA Vaccines and Therapeutics. Under Review.

14. **Coll De Peña, A.**; Gutterman-Johns, E.; Gautam, G.P.; Rutberg, J.; Ben Frej, M.; Mehta, D.R.; Shah, S.; Tripathi, A. Assessment of pDNA Isoforms using Microfluidic Electrophoresis. Under Review.

13. **Coll De Peña, A.**; Li, N.S.; Vaduva, M.; Goswami, S.; Tripathi, A. Empirical analysis and physics-informed neural network-based prediction of mRNA and dsRNA electrophoretic behavior. Under Review.

12. **Coll De Peña, A.**; Zimmer, D.; Gutterman-Johns, E.; Chen, N.; Tripathi, A.; Bailey-Hytholt, C.M. Electrophoretic Microfluidic Characterization of mRNA and pDNA Loaded Lipid Nanoparticles. Under Review.

11. **Coll De Peña, A.**; White, J.D.; Mehta, D.R.; Ben Frej, M.; Tripathi, A. Microfluidic AAV Purity Characterization: New insights into serotype and sample treatment variability. ACS Omega (accepted).

10. **Coll De Peña, A.**; Li, N.; Vaduva, M.; Bwanali, L.; Tripathi, A. A microfluidic electrophoretic dual dynamic staining method for the identification and relative quantitation of dsRNA contaminants in mRNA vaccines. Analyst, 2023, 148, 3758-3767.

9. Aibel, C.†; **Coll De Peña, A.**†; Tripathi, A. An Optimized CoBRA Method for the Microfluidic Electrophoresis Detection of Breast Cancer Associated *RASSF1* Methylation. BioTech, 2023, 12(1), 7. † Co-first authors, contributed equally.

8. **Coll De Peña, A.**; Masto, L.; Atwood, J.; Tripathi, A. Electrophoresis-Mediated Characterization of Full and Empty Adeno-Associated Virus Capsids. ACS Omega 2022, 7 (27), 23457-23466.

7. Hill, N.; **Coll De Peña, A.**; Miller, A.; Lapizco-Encinas, B.H. On the potential of microscale electrokinetic cascade devices. Electrophoresis, 2021, 42: 2474-2482.

6. Antunez-Vela, S.; Perez-Gonzalez, V.H.; **Coll De Peña, A.**; Lentz, C.J.; Lapizco-Encinas, B.H. Simultaneous Determination of Linear and Nonlinear Electrophoretic Mobilities of Cells and Microparticles. Analytical Chemistry 2020, 92, 22, 14885–14891.

5. **Coll De Peña, A.**; Hill, N.; Lapizco-Encinas, B.H. Determination of the Empirical Electrokinetic Equilibrium Condition of Microorganisms in Microfluidic Devices. Biosensors 2020, 10, 148.

4. Quevedo, D. F.; Lentz, C. J.; **Coll de Peña, A.**; Hernandez, Y.; Habibi, N.; Miki, R.; Lahann, J.; Lapizco-Encinas, B. H. Electrokinetic characterization of synthetic protein nanoparticles. Beilstein J. Nanotechnol. 2020, 11, 1556–1567.

3. **Coll De Peña, A.**; Miller, A.; Lentz, C.J.; Hill, N.; Parthasarathy, A.; Hudson, A.O.; Lapizco-Encinas, B.H. Creation of an electrokinetic characterization library for the detection and identification of biological cells. Anal. Bioanal. Chem. 2020, 412, 3935–3945.

2. **Coll De Peña, A.**; Mohd Redzuan, N.H.; Abajorga, M.K; Hill, N.; Thomas, J.A.; Lapizco-Encinas, B.H. Analysis of Bacteriophages with Insulator-Based Dielectrophoresis, Micromachines, 10(7) 450, 2019.

1. Thomas, J.A.; Benítez Quintana, A.D.; Bosch, M.A.; **Coll De Peña, A**.; Aguilera, E.; Coulibaly, A.; Wu, W.; Osier, M.V.; Hudson, A.O.; Weintraub, S.T.; Black, L.W. Identification of essential genes in the Salmonella phage SPN3US reveals novel insights into giant phage head structure and assembly. J. Virol. 2016, 90, 10284–10298.

Conference Presentations

33. **Coll De Peña, A.**; Tripathi, A. Rapid characterization of biopharmaceuticals using microfluidic electrophoresis, CE in the Biotechnology & Pharmaceutical Industries, Philadelphia, PA, 9/23. Oral presentation by **Coll De Peña, A.**

32. **Coll De Peña, A.**; Li, N.; Vaduva, M.; Bwanali, Lloyd; Tripathi, A. Detection and characterization of dsRNA contaminants in mRNA vaccines, BioProcess International, Boston, MA, 9/23. Oral presentation by **Coll De Peña, A.**

31. **Coll De Peña, A.**; White, J.D.; Mehta, D.R.; Ben Frej, M.; Tripathi, A. Rapid Characterization of Adeno-Associated Virus Samples: Full & Empty Capsids, Viral Proteins & ssDNA, Revvity 2023 Microfluidic CE Symposium, Boston, MA, 9/23. Oral presentation by **Coll De Peña, A.**

30. **Coll De Peña, A.**; Li, N.; Vaduva, M.; Bwanali, Lloyd; Tripathi, A. Microfluidic Electrophoresis-Based Detection of dsRNA Contaminants in mRNA Vaccines, Revvity 2023 Microfluidic CE Symposium, Boston, MA, 9/23. Oral presentation by **Coll De Peña, A.**

29. **Coll De Peña, A.**; Li, N.; Vaduva, M.; Bwanali, Lloyd; Tripathi, A. Development of a microfluidic electrophoresis dual dynamic staining method for the detection and assessment of dsRNA contaminants in mRNA vaccines, ACS Fall 2023, San Francisco, CA, 8/23. Oral presentation by **Coll De Peña, A.**

28. **Coll De Peña, A.**; Li, N.; Vaduva, M.; Tripathi, A. Understanding the electrokinetic mobility differences between dsRNA and mRNA within microfluidic systems, ACS NERM, Boston, MA, 6/23. Oral presentation by **Coll De Peña, A.**

27. Zimmer, D.; **Coll De Peña, A.**; Tripathi, A.; Bailey-Hytholt, C.M. Assessment of lipid nanoparticles: Developing a method to quantify encapsulated nucleic acid per particle, ACS NERM, Boston, MA, 6/23. Oral presentation by Zimmer, D.

26. Tripathi, A.; **Coll De Peña, A.** Interactive Learning Session: Understanding How Automation Quality Control Techniques Can Be Used to Successfully Design mRNA Workshop, mRNA Quality Control & Comparability Summit, Boston, MA, 6/23. Workshop by Tripathi, A.

25. **Coll De Peña, A.**; Tripathi, A. High Throughput Microfluidic mRNA Purity Assessment Method to Support mRNA Production, mRNA Quality Control & Comparability Summit, Boston, MA, 6/23. Oral presentation by **Coll De Peña, A.**

24. **Coll De Peña, A.**; Li, N.; Vaduva, M.; Bwanali, Lloyd; Tripathi, A. Microfluidic electrophoresis-based detection of dsRNA contaminants in mRNA vaccines, PEGS, Boston, MA, 5/23. Oral presentation by **Coll De Peña, A.**

23. Zimmer, D.; **Coll De Peña, A.**; Bailey-Hytholt, C.M. Developing a biomimetic model of placental drug transport, Biomedical Engineering Society: Advanced Bio-manufacturing Special Interest Group, Hyattsville, MD, 3/23. Poster by Zimmer, D.

22. **Coll De Peña, A.**; Li, N.; Vaduva, M.; Bwanali, L.; Tripathi, A. Microfluidic electrophoretic dual dynamic staining method for identification and relative quantification of dsRNA contaminants in mRNA vaccines, American Physician Scientists Association Brown Chapter Conference, Providence, RI, 3/23. Poster by Li, N.

21. **Coll De Peña, A.**; Chen, N.; Gutterman-Johns, E.; Zimmer, D.; Bailey-Hytholt, C.M.; Tripathi, A. High Throughput Microfluidic Purity Assessment Method for Gene Therapies, SLAS 2023 International Conference, San Diego, CA, 2/23. Poster by **Coll De Peña, A.**

20. **Coll De Peña, A.**; Tripathi, A. Microfluidic Electrophoretic Characterization of Full and Empty Adeno-Associated Virus Capsids. Exploring Characterization Methods: How to Completely Characterize Empty & Full Virus Particles Workshop, Gene Therapy Analytical Development Summit, Boston, MA, 11/22. Workshop speakers and panelists: **Coll De Peña, A.**; Draper, B.; Bechill, J.; Katsikis, G.; Wang, W.

19. **Coll De Peña, A.**; Tripathi, A. Electrophoresis-Mediated Characterization of Full and Empty Adeno-Associated Virus Capsids, The 2022 SACNAS National Diversity in STEM Conference, San Juan, Puerto Rico 10/22. Oral presentation by **Coll De Peña, A.**

18. **Coll De Peña, A.**; Aibel, C.; Tripathi, A. Microfluidic electrophoresis detection of breast cancer associated *RASSF1* methylation, Recent Advances in Biomedical Diagnostic Platforms Workshop, Institute of Medical Sciences in Mongolia (Online), 9/22. Oral presentation by **Coll De Peña, A.**

17. Li, N.; **Coll De Peña, A.**; Vaduva, M.; Tripathi, A.; Detection and Characterization of dsRNA contaminant in mRNA vaccines using microfluidic electrophoresis, Research Symposium at Brown University, Providence, RI, 8/22. Poster by Li, N.

16. **Coll De Peña, A.**; Tripathi. A. Microfluidic electrophoresis mediated characterization of Adeno-Associated Virus purity, New England Regional SACNAS Meeting, Providence RI, 6/22. Oral presentation by **Coll De Peña, A.**

15. **Coll De Peña, A.**; Masto, L.; Tripathi. A. Microfluidic electrophoretic assessment of Adeno-Associated Virus purity, PEGS Boston, Boston, MA, 5/22. Oral presentation by **Coll De Peña, A.**

14. **Coll De Peña, A.**; Masto, L.; Tripathi, A. Microfluidic Electrophoresis-Based Assessment of Adeno-Associated Virus Purity in Terms of Full and Empty Capsids, PEGS Boston, Boston, MA, 5/22. Poster by **Coll De Peña, A.**

13. **Coll De Peña, A.**; Tripathi, A. Biomedical Engineering and Microfluidics at Virtual Annual Research Week, Universidad Nacional Pedro Henríquez Ureña, Dominican Republic (online), 7/21. Oral presentation by **Coll De Peña, A.**

12. Antunez-Vela, S.; Perez-Gonzalez, V.H.; **Coll De Peña, A.**; Lentz, C.J.; Lapizco-Encinas, B.H. Characterizing unusual particle electromigration effects in a straight PDMS channel at SciX 2020 Virtual Meeting, 10/20. Poster by Antunez-Vela, S.

11. Miller, A.; Hill, N.; **Coll De Peña, A.**; Lapizco-Encinas, B.H. Electrokinetic separation of cells and polystyrene particles in an insulator-based device, Annual AES Meeting at SciX 2020 Virtual Meeting, 10/20. Poster by Miller, A.

10. Hill, N.; **Coll De Peña, A.**; Miller, A.; Lapizco-Encinas, B.H.; Microparticle filtration and separation using cascade devices, Annual AES Meeting at SciX 2020 Virtual Meeting, 10/20. Poster by Hill, N.

9. **Coll De Peña, A.**; Mohd Redzuan, N.H.; Abajorga, M.K; Hill, N.; Thomas, J.A.; Lapizco-Encinas, B.H.; Electrokinetic assessment of bacteriophage virus, Annual AES Meeting at Scix, Palm Springs CA, 10/19. Oral Presentation by **Coll De Peña, A.**

8. **Coll De Peña, A.;** Parthasarathy, A.; Hudson, A.O.; Lapizco-Encinas, B.H.; Insulator-based dielectrophoresis to characterize the electrokinetic behavior of bacterial and yeast cells and create a library, Annual AES Meeting at Scix, Palm Springs CA, 10/19. Poster by **Coll De Peña, A.**

7. **Coll De Peña, A.;** Hill, N.; Lapizco-Encinas, B.H.; Multi-stage electrokinetic microsystem for the purification and assessment of microbes, Annual AES Meeting at Scix, Palm Springs CA, 10/19. Poster by **Coll De Peña, A.**

6. **Coll De Peña, A**.; Lapizco-Encinas, B.H.; Insulator-based dielectrophoresis to characterize the electrokinetic behavior of bacterial cells and create a library, Scix, Atlanta GA, 10/18. Oral Presentation by **Coll De Peña, A.**

5. **Coll De Peña, A.;** Mohd Redzuan, N.H.; Thomas, J.A.; Lapizco-Encinas, B.H.; Insulator-based dielectrophoresis to purify and enrich bacteriophages, Annual AES Meeting at Scix, Atlanta GA, 10/18. Poster by **Coll De Peña, A.**

4. **Coll De Peña, A.;** Mohd Redzuan, N.H.; Thomas, J.A.; Lapizco-Encinas, B.H.; Insulator-based dielectrophoresis to purify and enrich bacteriophages, RIT Undergraduate Research Symposium, Rochester NY, 8/18. Poster by **Coll De Peña, A.**

3. Weintraub, S.T.; **Coll, A.**, Bosch, M.; Hajala, K.; Pardo, S., Molleur, D.; Hardies, S. C.; Black, L. W.; Thomas, J.A.; Novel Genetic Determinants for Complex Head Composition in a Giant Salmonella Phage, American Society for Mass Spectrometry Annual Conference, San Antonio, Texas, USA, 6/16. Poster by Weintraub, S.T.

2. Thomas, J.A.; Adams, L.E.; Desmond, M.I.; **Coll De Peña, A.**; Bosch, M.; Benitez, D.; Hardies, S.C.; Weintraub, S.T.; Black, L.W.; Revisiting the classics: *Salmonella* phage SPN3US as a model for understanding giant myovirus structure and life-cycle, Molecular Genetics of Bacteria and Phages Meeting, Wisconsin, Madison USA, 8/16. Poster by Thomas, J.A.

1. Thomas, J.A.; **Coll, A.**; Bosch, M.; Adams, L.; Benitez, D.; Aguilera, E.; Coulibaly, A.; Cheng, N.; Wu, W.; Steven, A.C.; Weintraub,S.T.; Hardies, S.C.; Black, L.W.; Exploiting Mutational Surrogacy to Study Head Morphogenesis of Giant PhiKZ-related Phages, Evergreen International Phage Meeting, Evergreen State College, Olympia, Washington USA, 10/15. Oral Presentation by Thomas, J.A.

Research Experience

Graduate Researcher September 2020-December 2023
Center for Biomedical Engineering, School of Engineering
Brown University, Providence, RI
Advisor: Professor Anubhav Tripathi, Ph.D.
Characterization of Nucleic Acid Delivery Vehicle Products using Microfluidic Concepts

- Developed a mathematical formulation and microfluidic methodology capable of assessing the purity of various adeno-associated virus (AAV) serotypes in terms of full capsids.
- Developed and optimized a workflow for the rapid determination of nucleic acid load in lipid nanoparticle (LNP) samples loaded with pDNA and mRNA.
- Investigated the electrokinetic behavior and differential biochemical responses of RNA molecules within microfluidic systems to inform the development of microfluidic-based analytical methods to characterize dsRNA impurities in mRNA vaccines.
- Developed a microfluidic electrophoresis method for the detection and characterization of the three predominant pDNA isoforms.
- Created microfluidic-based workflows for the screening of breast cancer and aneuploidy.
- Led a team of undergraduate students and collaborated with academia and industry to generate interdisciplinary solutions for analytical challenges of the biopharmaceutical industry.

Research Assistant August 2017-July 2020
Department of Biomedical Engineering
Rochester Institute of Technology, Rochester, NY
Advisor: Professor Blanca Lapizco-Encinas, Ph.D.

- Electrokinetically characterized, separated, and enriched microbes in microfluidic devices.
- Created an electrokinetic library to identify bacteriophages, yeast, and bacterial cells using microfluidic systems.

Undergraduate Researcher August 2016-August 2017
Thomas H. Gosnell School of Life Sciences, Biotechnology
Rochester Institute of Technology, Rochester, NY
Advisor: Professor André O. Hudson, Ph.D.

- Cloned and enzymatically characterized L,L-diaminopimelate aminotransferase derived from plant RNA using bacterial cultures.

Undergraduate Researcher August 2014-May 2016
Thomas H. Gosnell School of Life Sciences, Biotechnology
Rochester Institute of Technology, Rochester, NY
Advisor: Professor Julie A. Thomas, Ph.D.

- Assisted with establishing a novel system for the genetic analysis of large prokaryotic virus using bacteriophage SPN3US.

Research Intern Summers 2014, 2015
Institute of Molecular Biotechnology
Austria Academy of Sciences, Vienna, Austria
Advisor: Professor Josef Penninger, Ph.D.
Supervisors: Rezaul Karim, Ph.D., and Blanca Pranjic, Ph.D.

- Assisted with studying the role played by oncogene and tumor suppressor genes in the metastasis of cancer using *in vivo* and *in vitro* models.

Teaching and Mentorship

Research Mentor, Brown University May 2021-present

- Mentored eight Brown University undergraduate students from five different disciplines and one summer high school student.
- Taught and performed experiments one-on-one with each student, discussed the results and scientific insights resulting in two honors theses and eight publications.

Co-Instructor, Brown University January 2023-May 2023

- Phage Hunters II: Spring 2023.
- Worked on curricular development, planning the primary literature review portion of the course and curating the research articles that the students would be dissecting in class.
- Taught first year biology students how to read, interpret and present scientific findings the audiences with different scientific backgrounds.
- Taught one lecture per week and held office hours to guide students through their assignments.

Teaching Assistant, Brown University September 2021-December 2021

- Phage Hunters I: Fall 2021.
- Assisted students in the laboratory-intensive class during the isolation and characterization of bacteriophages.
- Evaluated the current class curriculum and proposed new assignments for the future terms.

Teaching Assistant, Rochester Institute of Technology August 2016 - December 2016

- Molecular Biology for Engineers: Fall 2016.
- Assisted students in the laboratory with their tissue cultures, helping them stay on track by monitoring their progress through the semester.

Teaching Assistant, Rochester Institute of Technology January 2016 - May 2016

- Phage Biology: Spring 2016
- Provided students with guidance with their research projects and helped them develop new laboratory skills in the field.

Honors and Awards

- CASSS Next Generation Investigator Award (2023).
- Recipient (Co-PI) of Brown University Data Science Initiative Seed Grant (2023).
- SLAS International Conference Tony B. Travel Award (2023).
- ACS Bridge Project: Career Kick-Starter Workshop Invited Attendee (2021).
- RIT Gosnell School of Life Science Research Scholar Award (2019).
- RIT Graduate Commencement Delegate of the College of Liberal Arts (2019).
- RIT Women in Science Student Travel Award (2019).
- National Science Foundation Travel Grant (2018).
- RIT Student-Faculty Research Grant (2014 and 2015).
- RIT Dean's List (2014 to 2019).

Leadership and Professional Development

Diversity and Inclusion Action Committee, CBME, Brown University 2022-present

Business Strategy Certificate, Harvard Business School Online 2023

Outreach Committee, Society of Women Engineers, Brown University 2021-2022

2023 New England Regional SACNAS Conference Organization Committee 2021-2022

Affiliations

American Chemical Society 2023-present
Member

Society for Advancement of Chicanos/Hispanics and Native Americans in Science 2021-2023
Member of the national organization and treasurer of the Brown University Chapter (2021-2023)

American Electrophoresis Society 2018-2019
Member

Delta Phi Alpha National Honorary German Fraternity, Tau Rho Chapter 2015-2019
Founding member, president (2015-2017, 2019) and vice-president (2017-2019)

Acknowledgments

There are so many people who have contributed to my scientific journey that I would like to thank, but I cannot trust my memory to remember all, so I will, to the best of my ability, highlight the ones without whom I would not be here today.

First, I would like to thank my Ph.D. advisor, Dr. Anubhav Tripathi, for pushing me and motivating me to accomplish goals beyond my imagination in what has been one of the most challenging yet rewarding experiences of my life. Always encouraging me to focus on finding solutions for real-life problems without losing sight of the fundamentals that drive them. I genuinely look forward to receiving your mentorship and guidance through the next stages of my life.

Next, I would like to thank the professors who have supported my progress through the main milestones of the degree: Dr. Eric Darling, Dr. Anita Shukla, Dr. Ian Wong, Dr. Robert Hurt, Dr. Somdatta Goswami, Dr. Christina Bailey-Hytholt, Dr. Kareen Coulombe, and Dr. Michelle Dawson. Similarly, I would like to acknowledge current and past Center for Biomedical Engineering members, particularly Dr. Celinda Kofron, Dr. Marissa Gray, Jess Bello, Carolyn Sherman, Lori Bassett, and John Lee, for their support. I also want to thank my fellow Center for Biomedical Engineering and SACNAS students, especially Adriana and Kimmy, for teaching me the importance of community, diversity, and inclusion.

I would also like to express my gratitude to my past research mentors who taught me the fundamentals of research and motivated me by example to pursue a Ph.D.: Dr. Blanca Lapizco-Encinas, Dr. André O. Hudson, Dr. Julie A. Thomas, Dr. Blanka Pranjic, and Dr. Karim Rezaul. I

would also like to thank my official and unofficial collaborators and mentors. Particularly Nicole Hill, Weixi Liu, Lloyd Bwanali, Menel Ben Frej, Jimmy White, Dipti Mehta, and Daniel Zimmer.

The next group I would like to thank is my Ph.D. lab mates from the Tripathi Lab, Lindsay, Kiara, Cel, Ramisa, JP, Duuluu, Jenna, Sarah, and Katie, for their company, support and collaboration, as well as the other members I was fortunate to overlap with during my time in the lab. I want to thank my incredible team of students, whose curiosity and engagement motivated me to persevere when our projects became challenging. While I was their research mentor, I learned just as much, if not more, from them: Claire, Lucy, Nina, Matei, Everett, Nicole, Johnny, Natalie, and Julie.

Last but definitely not least, I would like to thank the people behind the scenes who made all of this possible by providing me with their constant and unwavering support, my family. While I am unsure where to start because words cannot describe my gratitude, the first person I have to acknowledge is Tony, who has been there for me every step of the way from the day I submitted my Ph.D. application. Thank you for cheering me up when I needed to be cheered, for taking care of me when I did not have the time or energy to do it, and for being my partner across this journey and all our expeditions. Next, I want to thank my mom, Hinya, who, despite all adversity, managed to give my sister and me a life we could never repay her for. Thank you for being the strongest role model I could have asked for and for teaching me resilience. Thank you for always believing in me and encouraging me to follow my dreams even when you were never able to follow yours. I also want to acknowledge Ysabela, Ludovino, José Alejandro, Dino, and Bali for reminding me to take a break and enjoy life, mostly underwater. And, of course, my friends, both local and international. Thank you all for accompanying me through the easy and the hard times and always looking after me.

Contents

List of Figures

List of Tables

List of Supporting Information Figures

List of Supporting Information Tables

Chapter 1: Introduction

1.1 Motivation

Biopharmaceuticals have transformed the healthcare field by enabling the treatment of previously untreatable diseases[1] and decreasing the time from pathogen discovery to vaccine production.[2–4] However, the increased complexity of this modality over traditional chemical pharmaceuticals comes with significant analytical challenges (Figure 1.1). For instance, the molecular weight of a viral gene therapy such as adeno-associated virus is 27,777 times greater than that of aspirin, increasing the possible product distribution range by several orders of magnitude. Their increased molecular complexity makes studying their chemical and biochemical properties with the current mechanistic models not trivial. As a result, the lack of high sensitivity and resolution analytical methods makes it difficult to quantify undesired molecular variants,[5] which play an integral role in the safety and efficacy of biotherapeutics. Therefore, new experimental techniques are needed to generate data for novel quantification and predictive models for biopharmaceutical characterization. In addition, given the iterative nature of the production of biopharmaceuticals, novel analytical methods should have high-throughput capabilities to adequately support the analysis of the numerous samples generated during this process.

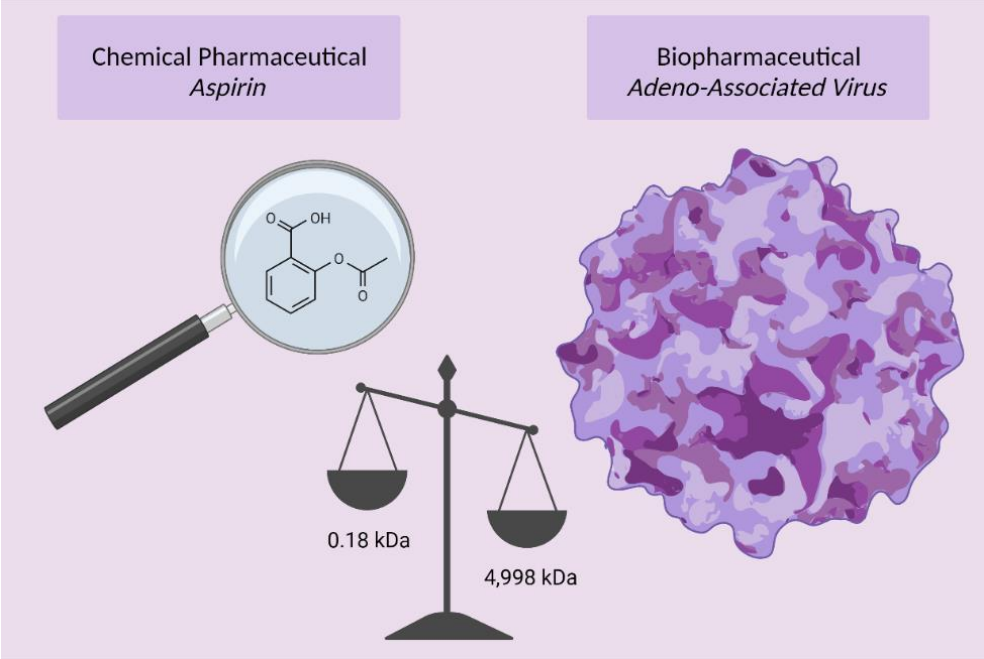

Figure 1.1: Schematic highlighting the increased complexity of biopharmaceuticals (*i.e.*, adeno-associated virus) over traditional chemical molecules (*i.e.*, aspirin).

Despite constant advancements in the field, traditional analytical characterization methods are severely hampered by time-limiting and personnel-intensive steps, leading to prohibitive costs and high turnaround times. The miniaturization of analytical systems has opened the doors to rapid, high-resolution, and high-throughput alternatives to traditional analytical methods. With its great automation compatibility and low volume requirements,[6] microfluidics presents a promising solution to the increased analytical burden of novel biopharmaceuticals. Thus, the dissertation focuses on deepening the understanding of molecular transport within microscale geometries to inform the development of rapid, high-throughput analytical techniques for analyzing biopharmaceuticals, particularly nucleic acid delivery vehicle products.

1.2 Biopharmaceuticals

Since their first introduction in 1982,[7,8] biopharmaceuticals have transformed the healthcare field by enabling the treatment of previously incurable diseases[1] and streamlining vaccine development.[4] Broadly defined as biological pharmaceuticals manufactured using biotechnology,[7]

they encompass recombinant proteins, and nucleic acid- and genetically engineered cell-based products.[9,10] With a growing need for personalized and rapidly produced drug products, the biopharmaceutical market has grown faster than the market for all drugs and is expected to keep growing.[9] Currently, the biopharmaceutical market is valued at USD 407 billion and is expected to reach USD 652 billion by 2028.[11] From January 2018 to June 2022 alone, 197 biopharmaceutical products received approval in the United States and Europe alone.[9] The COVID-19 mRNA vaccines, which were the first of their kind to receive Food and Drug Administration (FDA) approval, are clear examples of the potential of biopharmaceuticals, having gone from pathogen discovery to vaccine production at an unprecedented speed.[4] However, as the field grows exponentially, so does the need for streamlined alternatives to every step of their production workflow.

1.3 Nucleic Acid Delivery Vehicle Products

Among the numerous types of biopharmaceuticals, nucleic acid delivery vehicles stand out because of their ability to deliver genes into specific cells[12,13] through gene therapies and to manipulate gene expression and protein synthesis through nucleic acid-based vaccines and therapeutics.[14–16]

Genetics plays a role in most common diseases, either as a causative agent or as a facilitator, being attributed with most of the morbidity and mortality in developed regions, which stem from coronary heart disease, diabetes, different types of cancer, and depression.[17] In recent years, gene therapy has shown great promise in treating these genetic variations. Similarly, advances in genome sequencing, which have decreased cost and turnaround time, have supported the development of nucleic acid-based vaccines and therapeutics, which rely on this genetic information to design and synthesize potential products.[14,15,18]

While nucleic acid vaccines are usually delivered through non-viral systems such as lipid- and polymer-based, gene therapies routinely use viral and non-viral systems.[13] This work aims to address some of the critical analytical challenges associated with both types of systems. In addition, the different types of nucleic acids loaded onto these vehicles will also be studied here, particularly ssDNA, mRNA, ssRNA, dsRNA, and pDNA.

1.3.1 Non-Viral Delivery Vehicles

Non-viral delivery systems span both organic and inorganic materials. They are considered attractive delivery vehicles because of their low immunogenicity and enhanced cost-effectivity, particularly in comparison to the more challenging to manufacture cost-effectively viral vectors.[19] Some notable non-viral delivery systems that have obtained FDA approval include lipid nanoparticles (LNPs) as a vaccine against COVID-19[20] and as a therapy to treat polyneuropathy,[21] polymer-based nanoparticles for multiple sclerosis and cancers,[22] and metallic nanoparticles for treating anemia in patients with chronic kidney disease. While the Moderna (Spikevax) and Pfizer/BioNTech (Comirnaty) mRNA-LNP COVID-19 vaccines were the first mRNA-LNP vaccines to receive FDA approval, Patisiran was the first company to receive FDA approval for an LNP product, its siRNA-LNP therapy for the treatment of polyneuropathy.[19,21] Because of their wide range of applications, input nucleic acid load, and current clinical relevance, LNPs were selected for further exploration within microscale systems.

1.3.2 Viral Delivery Vehicles

Despite the potential of non-viral delivery systems, until recent advancements in lipid formulations, they have been significantly less efficient than viral systems,[23] leading to increased research efforts being dedicated to viral delivery systems over the years. A key advantage viral vectors hold over non-viral vectors, to which part of their increased efficiency can be attributed, is

their innate ability to inject genetic material into the target host cells.[24] The most commonly used viral delivery vehicles are retrovirus, adenovirus, adeno-associated virus (AAV), lentivirus, and herpes simplex.[24,25] Currently, adenovirus, lentivirus, and AAV have the most clinical trials in order of most trials.[24] Some characteristics of AAV that make it an interesting vector with great potential are that it does not cause any known human diseases and cannot replicate without a helper virus, making it a dependoparvovirus.[24] They are also considered the least immunogenic viral vector, showing decreased levels of vector-related toxicity.[26,27] To add to this, the simplicity of their genome and wide range of tropisms make it a highly versatile platform capable of targeting a wide range of diseases in numerous cell or tissue types. As such, we decided to study the electrokinetic behavior of AAV and use this knowledge to inform the development of analytical methods for its characterization.

1.4 Analytical Methods

Given the iterative nature of the production process of biopharmaceutical development, there has been added pressure to streamline the different aspects of process development to decrease the mounting cost pressure faced by the biopharmaceutical industry.[28] An essential aspect of process development is using analytical methods to optimize production and monitor product quality. Analytical tools are crucial to our way of life with a wide range of applications, including molecular diagnostics and health surveillance,[29–31] environmental monitoring,[32] quality control of food,[33,34] and, of greater interest to this work, ensuring the quality of biopharmaceutical products.[35–40] However, as the complexity of these novel drug products increases, so do the attributes and the need for platforms capable of analyzing multiple attributes simultaneously[41–44] to decrease the analytical burden of the process. However, despite being highly accurate[45] or having high sensitivities,[46,47] most of the analytical gold standard methods rely on analytical tools

that are not scalable and have high turnaround times or that trade-off scalability for resolution.[48] Miniaturization of analytical systems, such as the use of microfluidics, has opened the doors to rapid, high-resolution alternatives that can be automated and parallelized, streamlining and decreasing the footprint of the analytical systems.[49]

1.5 Microfluidic Systems

In the 1990s, as technology enabled the miniaturization of systems, the field of microfluidics emerged and has seen a continuous increase in interest ever since.[50] From droplet dispensers to lab-on-a-chip devices, microfluidics has opened the doors to entirely new problem-solving approaches. As a miniaturized technology, microfluidics refers to the field that manipulates microscale amounts (10^{-9} to 10^{-18} L) of fluids within systems with dimensions in the scale of microns.[6] Through the decrease in overall scales, microfluidics offers precise control of external forces, high-throughput compatibility, and rapid sample processing, making it capable of outperforming traditional technologies and often leading to lower-cost alternatives.[6,51] To obtain such fine control over the dynamics within the microsystem, forces must be applied to generate fluid and sample movement. These driving forces are usually pressure, electrokinetic, or magnetic. However, once movement is generated, other forces, such as drag, may be introduced into the system through changes in the fluid composition. Electrokinetics was selected as the preferred force to achieve sample separation during analysis based on its ability to generate controlled fluid and sample movement.

1.5.1 Forces within Microfluidic Systems

The miniaturization of fluidic devices enables the exploitation of behaviors that are not easily achievable at a larger scale, such as linear velocity profiles and homogenous distribution of high-magnitude electric fields across the system.[6,51] This thesis proposes using electrokinetics to

generate precise control over molecular transport and dynamics within microscale systems. Electrokinetics (EK) is the umbrella term for the various phenomena observed when an electric field is applied to a fluid. Only direct current (DC) electric fields and linear electrokinetics will be considered and discussed here.

When a surface is exposed to an aqueous solution, the ionization and adsorption of ions on the surface will give the surface an electrical charge.[52] To complement and balance the newly formed electrical charge at the surface, a thin layer of ions will aggregate near the surface, creating what is referred to as the electrical double layer (EDL) or the Debye layer[52,53] (Figure 1.2a). Generally, when an electric field is applied within a non-microscale channel, if the system is interrogated with microscale resolution, fluid movement will be observed near the solid surface in contact with the aqueous solution. However, as the probe moves closer to the center of the channel, fluid movement will approximate zero. However, when an electric field is applied within a microchannel at a microscale resolution, electroosmosis (EO), which is fluid movement resulting from the application of an electric field, will generate a linear velocity profile at a microscale resolution resulting from the proximity of the opposing EDL and the homogenous distribution of the electric field within the system (Figure 1.2a). The other major component of EK generally observed within microfluidic devices is electrophoresis (EP), defined as the movement of a charged particle when exposed to an electric field.[54] Depending on the electrical potential at the surface, or zeta potential, of the particle, which may be manipulated by changing the pH of the suspending media, particles will move in a different direction (Figure 1.2b).

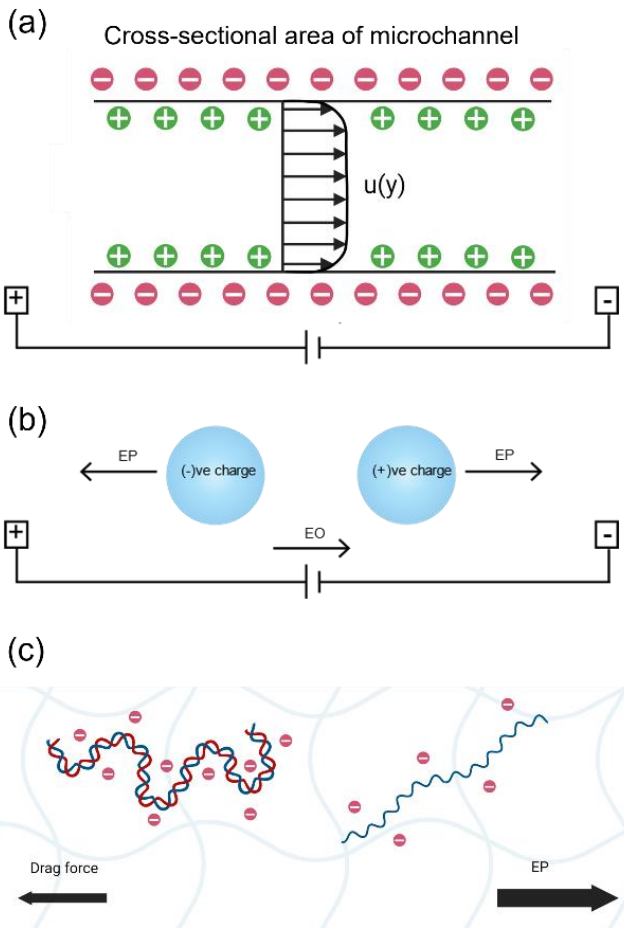

Figure 1.2: (a) Schematic representation of the electrical double layer. Schematic representation of (b) the electrophoretic (EP) and electroosmotic (EO) and (c) drag forces present when an electric field is applied within a microchannel.

Notwithstanding, while the presence of the EDL and EO results in a linear velocity profile, unlike EP, EO is independent of particle properties, causing its flow to potentially mask subtle differences between two particles. Thus, to mitigate the impacts of EO within the system, which may vary dramatically depending on media properties, a gel matrix, such as poly(*N,N*-dimethylacrylamide) or PDMA for short, can be used to coat the inner surfaces of the channels and decrease the extent of its EDL. The increase in viscosity of the suspending media will increase the effects of drag on the molecules as they migrate through the channel, making EP and drag the new predominant forces (Figure 1.2c). As both predominant forces are now directly dependent on particle properties,

more precise and targeted control of particle mobility can be achieved. As such, a wide range of particle properties such as size, shape, charge, and backbone flexibility have the ability to affect particle mobility, increasing the potential of the system to detect subtle differences between two similar particles.

Another advantage of having such fine control over the electrokinetic forces acting within the system is the ability to generate true plugs of sample. High-resolution analyses and separations can often be physically limited by the dimensions of the system since, usually, the shorter the separation channel, the lower the separation resolution. However, the ability to control both the fluid and sample dynamics with great accuracy allows for the creation of plugs containing finite volumes of sample (Figure 1.3) to be analyzed. By decreasing the size of the sample plug, the effects of the limited separation channel get mitigated as the dimensions of the channel will no longer be short relative to the volume being analyzed. As a result, the decreased dimensions, which require less sample volume and reagents and can yield improved resolution, make microfluidic platforms an attractive alternative to traditional biomolecular analytical tools.

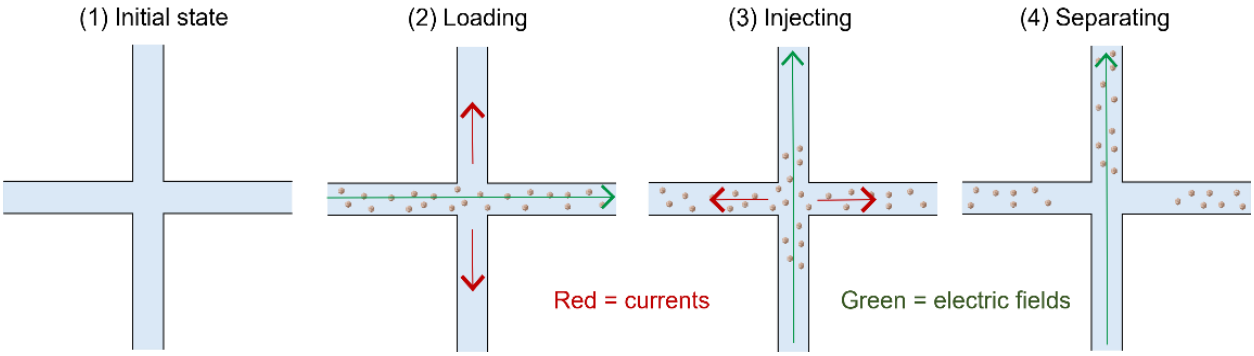

Figure 1.3: Schematic representation of the steps (voltages and currents) involved in the formation of a finite sample volume plug from an intersection within a microfluidic device.

1.6 Biopharmaceutical Application of Microfluidic Analysis

Currently, analytical gold standards for biopharmaceutical purity assessment tend to trade off throughput and resolution. For instance, analytical ultracentrifugation (AUC), the gold standard for assessing AAV composition, is not scalable and has a turnaround time of 2-6 hours.[45,55] In contrast, the preferred method for dsRNA contaminant detection in mRNA vaccines is dot blot;[46] however, while this method offers great sensitivity, it offers no resolution and cannot differentiate molecules of different lengths. Considering the need to analyze multiple attributes of the numerous samples generated during the iterative manufacturing of biopharmaceuticals, microfluidic systems offer a promising solution to the unmet analytical needs of the field that is both high-throughput and high-resolution.

Notwithstanding, despite their great potential, critical scientific challenges are associated with using microfluidics for biopharmaceutical characterization, stemming primarily from the limited intersection of these two modalities in the past. The major challenges that will be addressed in this work are a gap in the fundamental understanding of the molecular transport and separation mechanism of nucleic acid delivery vehicles in microscale geometries.

1.7 Summary and Research Overview

This thesis focuses on fundamental approaches to informing and developing rapid microfluidic electrophoresis analytical methods to streamline the purity assessment of biopharmaceuticals. To achieve this, great importance was placed on understanding the effect of the biophysical and biochemical properties of the particles on their electrokinetic mobility. A visual overview of this thesis and a description of how each chapter contributed to the focus of this dissertation is provided in Figure 1.4.

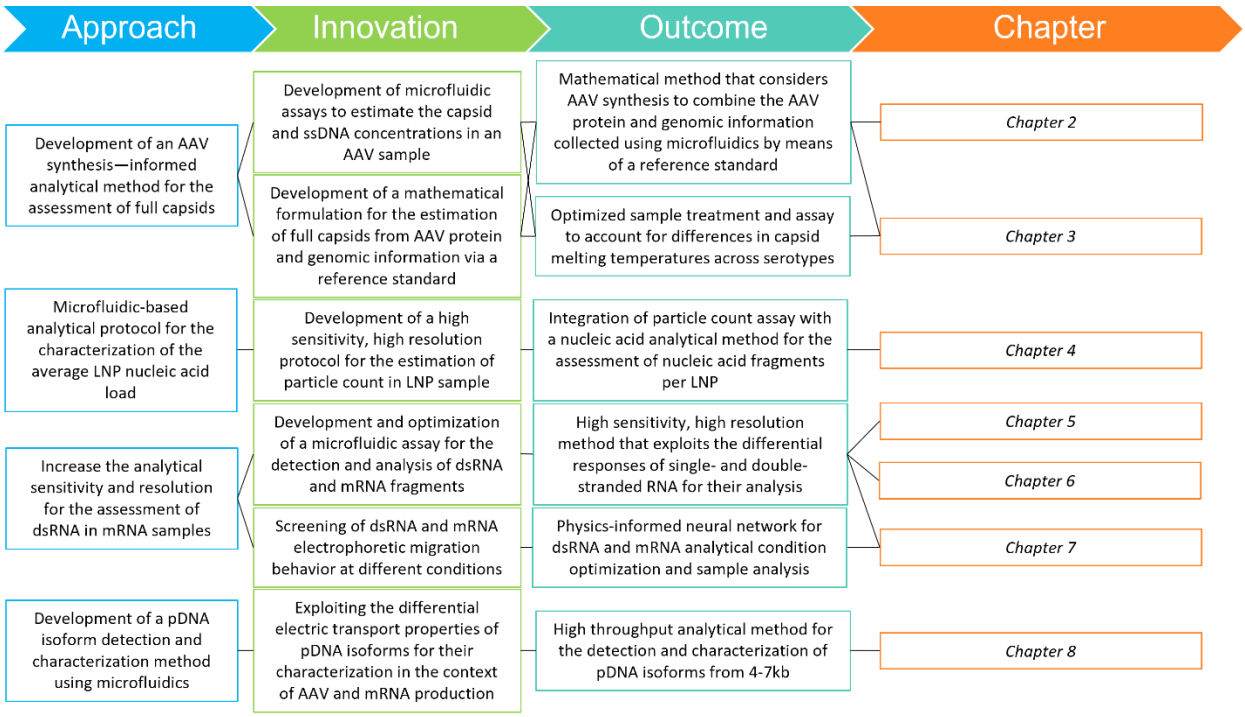

Figure 1.4: Outline of the contributions of this thesis to developing and streamlining analytical methods for biopharmaceutical purity assessment.

Chapter 2 describes the development of methodology for the assessment of AAV purity in terms of full and empty capsids, which combined experimental data with a mathematical formulation established for this method. For the experimental data, two independent microfluidic assays were developed and optimized, one to collect the genomic data and one to collect the capsid protein data of the AAV8 molecules. A reference standard of known composition was integrated into both workflows, and the mathematical formulation then used the reference standards to combine both assays and determine the genome-to-capsid ratio in each sample. This method was a significant improvement in terms of input sample volume and turnaround time over the existing methods.

Chapter 3 evaluates the biochemical and electrophoretic differences across AAV serotypes and uses this information to expand the range of the AAV purity assessment methodology presented in Chapter 2 to be compatible with multiple serotypes. By understanding how the differences in capsid melting temperatures affected the denaturing and digestion of each serotype, the method

was optimized for a broader range of melting temperatures. The findings in this chapter increased the understanding of the relationship between capsid melting temperature and serotype response, which informed the optimization of the AAV purity assessment method and increased its robustness.

Chapter 4 aims to understand how lipid-based systems, such as LNPs, behave within microscale geometries and how detergents can be used to analyze both their lipid and nucleic acid compositions. By systematically assessing the effects of different detergents on the lipid systems, optimal sample treatments were established for lipid and nucleic acid (pDNA and mRNA) visualization, and microfluidic analytical methods were developed and optimized for their analysis. Using the information yielded by each method, a reference standard was used to estimate the concentrations of each component of the LNP system and determine the average load of the vehicles for different formulations.

Chapter 5 begins our work toward understanding the biochemical and biophysical differences between dsRNA and mRNA. dsRNA is an immunostimulatory molecule generated as a by-product of the *in vitro* transcription of mRNA. Because of its effects on the immune response, it has to be monitored to ensure the safety and, possibly, efficacy of mRNA vaccines and therapeutics. A microfluidic electrophoresis method was developed to detect both types of molecules simultaneously. This assay was complemented by using two different fluorescent stains, a DNA and an RNA, and their differential fluorescent dynamic staining response was used for peak identification prior to size characterization. This work increased the understanding of the differential electrophoretic and dynamic staining responses of the two types of RNA molecules, providing an alternative to the traditional methods used to assess dsRNA impurities, which are incapable of assessing size.

Chapter 6 expands on the dsRNA and mRNA identification work by providing a more sensitive alternative to the dual dynamic staining method proposed in Chapter 5. A key limitation of the dual dynamic staining method was a relatively low maximum loading capacity of 13 ng/µL of mRNA, limiting the concentration percentage of dsRNA impurities that could be detected. To overcome this limitation, here, an enzymatic treatment using S1 nuclease is employed to fully digest the ssRNA and mRNA products present in the sample without affecting the dsRNA product's length. This treatment results in samples that only contain dsRNA products, which can be easily analyzed using microfluidics. By fully digesting the mRNA fragments, the maximum loading capacity of the system is increased to at least 200 ng/µL. In comparison, the sensitivity for dsRNA post-treatment is 0.32 ng/µL, enabling the detection of impurities at relative concentrations as low as 0.16%.

Chapter 7 builds upon the identification methods presented in Chapters 5 and 6 and aims to understand and model the electrophoretic behavior of dsRNA and mRNA in different analytical conditions for size assessment. As the scope of RNA vaccines and therapeutics expand, so do the lengths and compositions of the products. As such, it is essential to fundamentally understand the electrophoretic behavior of these molecules to rapidly inform the development of microfluidic analytical methods for their characterization. To this end, non-chemically and chemically (pseudouridine, like the mRNA COVID-19 vaccines) modified single- and double-stranded RNA molecules were analyzed at different PDMA concentrations in the 1-5% range. The empirical data was analyzed and compared to existing models in the literature. To complement the semi-qualitative analysis of the data, it was used to develop two types of artificial neural networks that could be used to predict the size of these molecules under different analytical conditions. It compared a data-driven model with a physics-informed model, and showed how integrating the

physical equations that govern the data significantly improved the prediction accuracy of the system. The ability of the physics-informed neural networks to predict with an average error of 0.77% the expected migration of RNA molecules shows the potential of machine learning to reduce the need for extensive experimentation and increase the understanding of electrophoretic mobility beyond RNA.

Chapter 8 shows how the high resolution of microfluidic electrophoresis can be exploited to rapidly differentiate the three predominant isoforms present in pDNA samples: supercoiled, linear, and open circular. In recent years, pDNA vectors have become of increasing interest due to their application as vaccines and as a precursor for mRNA vaccines and gene therapy products. In the case of mRNA vaccines, it is used supercoiled to amplify the gene of interest prior to its linearization and *in vitro* transcription. In the case of viral gene therapies, pDNA is commonly used supercoiled to transfect the cells used to produce the viruses. Depending on its intended application, the desired isoforms may be supercoiled or linear, with the possibility of any or a combination of the other isoforms being present as a contaminant. As such, analytical methods for assessing the isoforms are needed within different biopharmaceutical workflows and often at different stages of each workflow. To this end, a microfluidic electrophoresis method was developed to rapidly detect the three isoforms with only 10 μL of sample. In addition, the method can determine the size and concentration of the supercoiled and linear isoforms ranging from 0.1-15 ng/μL in 1 min/sample.

Chapter 9 provides concluding thoughts with a summary and description of the impact of the work presented, complemented by an overview of its limitations and future directions from this thesis.

Chapter 2: Electrophoresis-Mediated Characterization of Full and Empty Adeno-Associated Virus Capsids

---

Chapter 2 has been published as an original research article.

Coll De Peña, A.,[1] Masto, L.,[2] Atwood, J.,[3] Tripathi, A.,[1] *Electrophoresis-Mediated Characterization of Full and Empty Adeno-Associated Virus Capsids*, **ACS Omega**, 7 (27), 23457-23466, 2022.

[1]Center for Biomedical Engineering, School of Engineering, Brown University, Providence, RI, USA

[2]Division of Biology and Medicine, Brown University, Providence, RI, USA

[3]Applied Genomics, PerkinElmer, Hopkinton, MA, USA

2.1 Abstract

Adeno-associated virus (AAV) has shown great potential in gene therapy due to its low immunogenicity, lack of pathogenicity to humans, and ability to provide long-term gene expression in vivo. However, there is currently a need for fast, high-throughput characterization systems that require low volumes for the determination of its sample composition in terms of full and empty capsids since empty capsids are a natural by-product of AAV synthesis. To address this need, the following study proposes a high-throughput electrophoresis-mediated microfluidics approach that is independent of sample input concentration to estimate the composition of a given sample by combining its protein and ssDNA information relative to a standard. Using this novel approach, we were able to estimate the percentage of full capsids of six AAV8 samples with an average deviation from the actual percentage of 4%. The experiments used for these estimations were conducted with samples of varying percentages of full capsids (21–75%) and varying concentrations (5×10^{11}–1×10^{12} VP/mL) with a total volume requirement of 3–10 μL for triplicate analysis of the sample. This method offers a rapid way to evaluate the quality and purity of AAV products. We believe that our method addresses the critical need as recognized by the gene and molecular therapy community.

2.2 Introduction

Genetics is intrinsically involved in most common diseases, either as a causative agent or as a facilitator. While not generally associated with each other, the diseases responsible for most of the morbidity and mortality in developed regions, such as coronary heart disease, diabetes, different types of cancer, and depression, have, in fact, a significant genetic component.[17] Some of the most common variations in our DNA that result in genetic disorders include single-gene disorders, chromosomal imbalances, and epigenetics.[56] Consequently, for the past few decades, great importance has been given to advancements in human genome discoveries and the development of sequencing technologies due to their potential to improve health and prevent diseases.[57]

To address these genetic variations, in particular single gene disorders, significant improvements have been made to gene therapy since its early conceptualization over 40 years ago. Improvements range from the first publications of nonviral gene therapy in the 1980s to the successful use of viral and nonviral gene therapies in the treatment of cardiovascular diseases, autoimmune disorders, obesity, diabetes, and central nervous system disorders, among others.[58] Among the current gene therapy vectors, adeno-associated virus (AAV) has become of particular interest as it has been demonstrated to provide successful, long-term gene transfer *in vivo*.[59] To date, there are two FDA-approved AAV gene therapies: Luxturna for the treatment of a rare inherited retinal dystrophy, and Zolgesma for the treatment of spinal muscular atrophy.[60,61] In addition, dozens of treatments are currently in clinical trial, and another therapy for the treatment of adenosine deaminase-severe combined immunodeficiency, Strimvelis, has been approved by the European Medicines Agency.[61]

AAV is a single-stranded DNA (ssDNA) nonenveloped virus that belongs to the parvovirus family and measures approximately 25 nm in diameter.[62,63] There are currently nine serotypes of AAV, each with slightly different tropisms which include retina, lung, muscle, liver, and brain cells. The

virus is only composed of protein and DNA and has three repeating capsid proteins, VP1, VP2, and VP3, at an expected ratio of 1:1:10, which may vary across serotypes and even within the particles of a given batch.[64,65] It is able to infect both dividing and nondividing cells depending on the tropism of the given serotype.[66]

The simplicity of the AAV genome makes recombinant AAV (rAAV) a great candidate for gene delivery. While AAV generally requires a helper virus, such as an adenovirus or herpes simplex virus, to successfully replicate in human cells, a key change in the development of its design was the use of plasmids for rAAV production. The most common method is the use of the triple transfection of HEK293 cells, where one plasmid includes the gene of interest, another the helper genes, and a third the packaging genes.[67,68] The integration of the helper virus functional genes allows for the helper virus-independent manufacturing of recombinant AAV vectors. Additionally, while its restricted packaging limit of 4.7 kb has previously limited its scope, recent studies have been aimed at exploring the potential and integrity of oversized rAAV vectors.[69] Another recent innovation in the next generation of AAV vectors is the production of capsid-modified, genome-modified, and both capsid- and genome-modified vectors to increase their efficiencies at reduced doses.[70] Despite its great potential, there are still several limitations that prevent the safe and widespread use of this therapy, including the large-scale manufacturing of high-quality rAAV vectors. In addition, a streamlined quality control system for the rapid characterization of viable particles has not been developed. As a result, there is currently a need for rapid and robust assays for rAAV testing for human gene therapy.[71–73] The natural byproducts of AAV synthesis are capsids that have not been packaged with DNA, also referred to as empty capsids, and capsids with undesired loading of either fragmented ssDNA, referred to as partially full, or host DNA. It is not completely clear how the presence of empty capsids impacts the therapeutic properties of rAAV,

but their removal has shown an increase in transgene expression.[74] They have been shown to inhibit the transduction of target cells by exacerbating the immune response to rAAV and by competing with full capsids for cell binding sites, but they have also demonstrated beneficial effects by acting as decoys, which reduce the neutralization of full AAV vectors.[75] Regardless of their impact on therapeutic properties, it is important to have effective methods to differentiate full capsids from empty capsids. Despite the purification steps that have been integrated into the rAAV production chain, empty particles can still account for most of the particles in a batch.[76] Hence, quality control methods must be designed to monitor the production and purification of AAV suspensions.

Currently, several methods have been utilized in the purification process in attempt to determine the capsid content ratio (full vs empty capsids), as summarized in Table 2.1. Some of the most prevalent include anion-exchange chromatography (AEX), optical density (OD), transmission electron microscopy (TEM), charge detection mass spectrometry (CDMS), size-exclusion chromatography multiangle light scattering (SEC-MALS), enzyme-linked immunosorbent assay (ELISA) in tandem with quantitative polymerase chain reaction (qPCR), and analytical ultracentrifugation (AUC), in no particular order.[55,77,78]

Table 2.1: Comparison of Analytical Methods Used to Differentiate between Empty and Full AAV Capsids

Analytical Methods	Fundamental Basis	Characteristics
Anion exchange chromatography	Surface charge density	High accuracy, complex workflow, medium turnaround time (30 min/sample), difficult to establish optimal method applicable to all rAAV serotypes
Optical Density	UV absorbance (density analysis)	Fast turnaround time (15 min/sample), requires highly purified AAV to minimize interference with UV absorbance, varying concentration requirements (5×10^{11}–1×10^{13} GC/mL)
Transmission electron microscopy	Image analysis	Direct characterization method, statistically small sample image size, high coefficient of variation, time-consuming (6 h/sample)
Charge detection mass spectrometry	Mass to charge ratio	High accuracy, requires extensive preparation, time-consuming (2 h/sample), low throughput, not easily accessible
Size-exclusion chromatography, multiangle light scattering	Size exclusion and static light scattering	High accuracy, medium turnaround time (20–30 min/sample) requires long column equilibration times, relatively high volumes of sample (30–50 µL per run), high concentration requirements (1×10^{13} VP/mL)
ELISA + qPCR	Antibody specificity to full and empty capsids, used in tandem with qPCR	Expensive, not scalable, lacks accuracy and precision, compounded error
Analytical ultracentrifugation	Separates capsid sedimentation rate of particles (buoyant densities)	High accuracy, large volume of samples (300–400 µL), not scalable, time-consuming (6 h/sample)
LabChip electrophoresis (Our method)	Total protein/ssDNA ratio	Scalable, fast turnaround time (6–15 min/sample), high throughput, requires low volumes (3–10 µL) and concentrations (>1×10^{11} GC/mL)

Ion-exchange-based chromatography techniques, especially AEX, are often used in series and allow for enrichment of full capsids via the use of ion exchange resins, membranes, and monolithic columns; however, this approach can be time-consuming, requiring an analysis time of 30 min per sample and column equilibration between injections.[79–81] Although this approach remains time-consuming and each serotype requires condition optimization, it must be noted that recently

several groups have been working on the optimization of the technique across different serotypes.[81,82] OD, or UV absorbance spectroscopy, utilizes the extinction coefficient (an indicator of light absorbance at a given wavelength) of capsid proteins and DNA, and it estimates the ratio of capsid particle to vector genome based upon the absorbance ratio.[83] OD analysis time can be as low as 15 min/sample; however, it has very high purification requirements to prevent interference with the UV absorbance. Next is TEM which, when combined with a deep learning algorithm, allows for the visualization of the overall morphology of AAV samples via capsid staining. Those which are empty exhibit a stained electron-dense core, while those which are full exhibit a bright core.[84] CDMS has been used to differentiate between full, partially full, and empty capsids, but to achieve this, a complex and time-consuming workflow needs to be followed, and access to this technology remains limited.[85] SEC-MALS is another promising technique for the determination of full and empty capsids, which uses a form of static light scattering to determine the size and weight of the AAV molecules post size-exclusion. While this approach may provide rapid and highly accurate estimations, the process requires extensive SEC column calibration and relatively high volumes of sample (~50 μL) per individual run. An alternative approach exploits the genome to capsid ratio and requires the simultaneous use of ELISA to determine the capsid protein titer, in tandem with qPCR, to determine viral genome (ssDNA) titer. The empty particle content is determined by subtracting the viral genome titer from the capsid titer. This method risks compounded error from the utilization of two independent systems and assays and has only been optimized for certain serotypes.[86] AUC typically requires the samples to have a normalized concentration and requires extensive data acquisition (~60 scans) and data processing.[45] The robustness of AUC makes it the quantitative golden standard in the field for AAV quantification despite its high volume, concentration, and time requirements. Note that these are some of the

most established techniques currently available for capsid packaging characterization, but other novel approaches have been gaining traction and have shown great promise as alternative analytical platforms such as mass photometry, capillary isoelectric focusing, and Stunner (UV260/280).[40,87,88]

Despite the great potential of each of these techniques, as highlighted above, they are limited by their requirement of significant sample preparation or processing times.[55,81,85,89] In fact, sample analysis times for these techniques can range between 15 min (OD) and 6 h (AUC and TEM) per sample, depending on the technique, and require extensive resources, which significantly limits the number of samples that can be processed in a day, making most approaches low throughput.[55] In addition, it must be noted that oftentimes the reported analytical times associated with each method excludes the time of sample preparation, whereas the turnaround time that is reported for our proposed method combines the two.

Thus, the goal of this study is to develop a microfluidic technique that mathematically determines the ratio of full to empty particles in an unknown sample by using the genome to protein ratio of the unknown sample normalized to a standard. This approach is largely based on the fact that during AAV production capsids are synthesized prior to being packaged with their genome. Therefore, for a single serotype production, capsids will contain the same amount of protein regardless of their packaging, as represented in Figure S1. In other words, regardless of the presence and nature of genetic material within the capsids, the protein profiles of full and empty capsids should be the same. Please note that for the purpose of this study only full and empty capsids will be considered since together they represent >97% of the total AAV sample, especially in purified samples such as the ones used in this study.[55]

In the context of differentiating between full and empty rAAV capsids for gene therapy, microfluidic techniques can be utilized to characterize its capsids based on the genome to protein ratio of a given sample. Here, the goal is to use microfluidic electrophoresis that can quantify the presence of AAV single-stranded genomic DNA within full rAAV capsids via the application of voltages to digested viral capsids. This study aims to evaluate the ssDNA content in full capsids in tandem with the protein content of full and empty capsids to determine the percentage of full AAV capsids in a given sample of rAAV relative to a standard with a known fraction of full capsids, as highlighted in Figure 2.1. By utilizing microfluidic technology to create a reliable, high-throughput, and efficient composition assessment method for AAV capsid differentiation, this study will aim to make AAV an even more attractive candidate for gene therapy and allow for streamlined identification of successful versus unsuccessful insertion of genetic material.

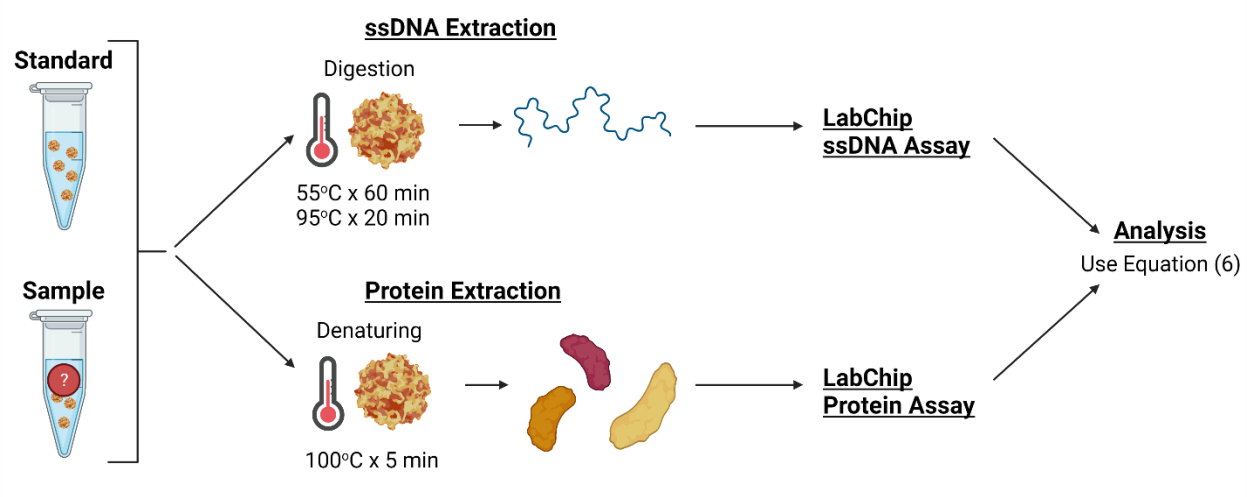

Figure 2.1: Workflow for AAV sample capsid content analysis. Here, the samples are mixed in a 384-well plate or in PCR tubes, heated following their respective protocols, and transferred onto the LabChip System. The sipper chip in the LabChip platform gets prepared following the protocol with a gel-dye matrix and a marker, together with the provided buffer and ladder tubes. Once the samples and chip are ready, the appropriate assay gets selected on the system, and the results get analyzed using the provided mathematical approach (eq 2.6).

2.3 Materials and Methods

2.3.1 Samples and Sample Preparation

The AAV8 full and empty reference standards were purchased from Vigene Biosciences (Vigene Biosciences, Rockville, MD). The full reference standard had a titer of 7.97×10^{11} genome copies (GC)/mL at a full fraction of 75%, which is the equivalent to 1.07×10^{12} viral particles (VP)/mL. The latter was estimated by accounting for the empty particles, which would not contribute to the genome copies. On the other hand, the empty reference standard was purchased with a titer of 1.44×10^{12} VP/mL at an empty fraction of 96%. Since this value includes both full and empty particles, the titer of only empty particles in this sample was estimated to be 1.38×10^{12} VP/mL. In the case of both the full and the empty reference standards, per the certificate of analysis issued by the provider (Vigene Biosciences), the quality control was conducted via SYBR Green qPCR and ELISA, combined with TEM to determine the percentage of full capsids in the sample. While the vendor did not provide a standard deviation for their estimations, a 10% deviation will be assumed, when needed, for statistical analysis, which may be consistent with the analytical tools used for their capsid ratio estimation. For the experiments comparing full to empty samples, the stock samples were used; however, the ones that used different percentages of full capsids were prepared accounting for the empty particles in the full reference standard and the full particles in the empty reference standard. Furthermore, samples were stored in single-use aliquots at −80 °C to ensure sample stability and degradation did not affect the outcomes of the experiments.

2.3.2 Methods

In this study, the GX Touch II LabChip system (PerkinElmer, Waltham, MA) was used to characterize the AAV samples. For the capsid protein experiments, the standard protocol of the LabChip ProteinExpress assay was used, as seen in Figure 2.1. Specifically, the high sensitivity

protocol was followed, which requires 5 µL of sample, using the optional reducing buffer. Each sample was analyzed three times, and each time 20 nL was removed from the well plate onto the detection chip. For the ssDNA analysis, a custom ssDNA chip assay was used (PerkinElmer). First, the AAV samples were digested 1:1 (5 µL of AAV and 5 µL of the digestion mixture) with a proteinase K mixture (10 µL of proteinase K were diluted with 90 µL of 2 M urea) for 60 min at 55 °C, followed by a proteinase K deactivation for 20 min at 95 °C, as seen in Figure 2.1. Here, it must be noted that while we are focusing on the detection of ssDNA, AAV could also contain self-complementary DNA, which may also be detectable with the simple insertion of an intercalating dye in the microchip gel formulation. Lastly, the samples were analyzed with a customized ssDNA assay script to amplify the signal and allow for the detection of samples with lower genomic content (PerkinElmer). Each sample was analyzed three times, and each time 20 nL was removed from the well plate onto the detection chip. Note that while 5 µL of sample was used for each assay, if needed, this volume could be reduced significantly without interfering with the assay. Lastly, the statistical analysis for this study was conducted using GraphPad Prism 9, and the figures were made using GraphPad and/or BioRender.com.

2.3.3 Mathematical Formulation to Estimate the Percentage of Full Capsids

The empty, full, and partial AAV capsids have icosahedral symmetry. They are assembled from viral proteins (VP1, VP2, and VP3) and genomic material (ssDNA). To fully characterize the AAV samples, quantitation of both proteins and ssDNA is needed. Before we describe our experimental findings, the following section describes a mathematical formulation to relate protein and DNA concentrations in the samples and an AAV standard.

Here, we describe a new method for the determination of the percentage of full capsids in an AAV sample. The method requires only a few microliters (3–10 µL) of a standard (S) with a known

fraction of full capsids. Let us denote the total number of AAV capsids in the sample and standard by N and N_s, respectively. Hence, the number of full and empty capsids can be described using the following equations

$$N = N(f) + N(e) \tag{2.1}$$

$$N_s = N_s(f) + N_s(e) \tag{2.2}$$

Here, f and e refer to full and empty, respectively. Note that the number of partially filled AAVs can be included in this analysis, but since their concentration is below our limit of detection (LOD) and their percentage negligible (<3%) in purified samples such as these, it was excluded from the analysis for simplicity. Since the subunits (VP1–3) of the AAV capsids are the same for both full and empty capsids, it is safe to assume that each capsid is composed of α μg of total proteins. Hence, the ratio of the concentration (c) of proteins in the sample and standard can be expressed as

$$R_p = \frac{c(protein)}{c_s(protein)} = \frac{\alpha N}{\alpha N_s} = \frac{N}{N_s} \tag{2.3}$$

Similarly, we assume a single ssDNA insert per full AAV capsid to obtain the following ratio for the concentration of ssDNA in the sample and standard:

$$R_{DNA} = \frac{c(ssDNA)}{c_s(ssDNA)} = \frac{N(f)}{N_s(f)} \tag{2.4}$$

Since the standard comes with a known percent of full capsids, β_s, which can be defined as

$$\beta_s = \frac{N_s(f)}{N_s(f)+N_s(e)}, \tag{2.5}$$

we obtain the fraction of full capsids, β, using the following relation:

$$\beta = \frac{N(f)}{N(f)+N(e)} = \frac{\beta_s R_{DNA}}{R_p} \quad (2.6)$$

Hence, we obtain the percentage of the full AAV estimate by measurement of concentrations of total protein and ssDNA using microfluidics electrophoresis with samples and a standard. It is important to note that our method is independent of the total capsid concentrations in samples or standards. This is often the most significant limitation for other techniques summarized in Table 2.1. The error in the estimated percentage of full AAV as determined by eq 2.6 only arises from the concentration ratio accuracy errors in the electrophoresis method, not from the resolution accuracy of the method. In other words, rather than obtaining the total protein concentration of a sample from the protein assay (full and empty) and subtracting the concentration obtained from the DNA assay (only full), which would have a compounded error rate from the use of two different assays, the proposed method uses a sample, or standard, of known concentration to normalize both assays. As an example, using the 75% full AAV8 sample, we can normalize the assay for each run to reduce the error introduced by the use of multiple assays and the difference in units between the protein and DNA areas. Another key advantage of the ratio measurements is that as long as all samples within the assay are treated in the same way (diluted by the same factor or heated for the same time) the proposed method is independent of the concentration of the sample since the protein area will account for a difference in concentration.

2.4 Results and Discussion

2.4.1 Capsid Protein Profile Characterization

The first step in the validation of the mathematical formulation to estimate the percentage of full capsids in a sample was to independently assess the protein and DNA profiles of AAV8 full (75%) and empty (96%, or 4% full) reference standards. At this stage, it was of particular interest to

confirm that the amount of protein in the full and empty samples at a given concentration of capsids was the same. To do this, the AAV standards were denatured using a reducing buffer containing DTT (refer to Figure 2.1 and Section 2.3.2 Methods). The concentration of the empty reference standard was adjusted to match the concentration of the full reference standard (1.07×10^{12} VP/mL) by diluting it with 1X PBS. The dilution was conducted keeping both the viral concentration and the salt composition, which can affect labeling efficiency and sample conductivity, constant.

On the basis of the electropherogram shown in Figure 2.2a, and as seen in the summarized corresponding areas under the curve for VP1, VP2, and VP3 (Figure 2.2b), there was no statistical difference between the full and empty samples VP peaks. Moreover, on the basis of the literature results and normalizing it to their respective molecular weights, VP1 is expected to represent 11.3% of the total capsid area, VP2 9.5%, and VP3 79.2%. In our analysis, the full sample VP1 represents 6.7%, 11.3% for VP2, and 82.1% for VP3. Similarly, the empty sample VP1 represents 6.3%, VP2 11.6%, and VP3 82.1%. The presence of these VP ratios that differ from the expected 1:1:10 ratio highlights the stochastic nature of the VP capsid protein ratios. Therefore, to mitigate the impact of the varying VP ratios across samples and even within a given batch, we decided to integrate the area under the three VP peaks to have a method that is independent of the VP ratios.

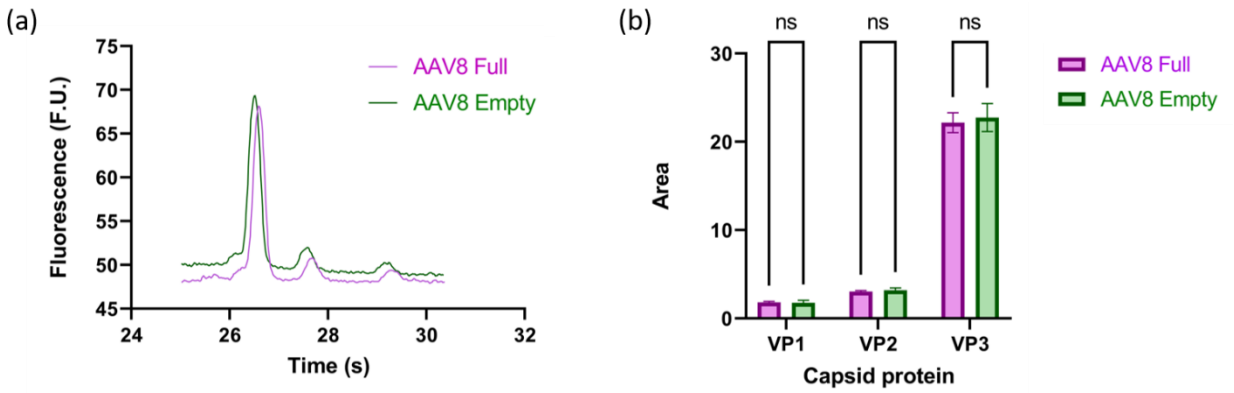

Figure 2.2: Analysis of capsid proteins from full (purple) and empty (green) AAV8 capsids. (a) Electropherogram representative of VP3, VP2, and VP1 capsid proteins, from lowest to highest molecular weight (left to right), of both full and empty samples. (b) Summarized area under the curve for each main protein peak, showing the statistical difference between the VP peaks yielded by the full and empty capsids.

Moreover, if the sample and standard (usually provided) concentrations are known, these values can be input as the concentration in a modified eq 2.3 instead of the protein area obtained in the Protein Express assay detailed above (refer to eqs 2.7 and 2.8). However, if the concentration is unknown, this approach can be used not only to estimate the percentage of full capsids in the sample but also to estimate the total concentration of the sample, as highlighted in Table 2.2. Using the area under the curve of the VP protein peaks of the standard and of the sample, as well as the concentration of the standard, we were able to estimate the total sample concentration with an error rate of 1–16% and an average error rate of 6%. Moreover, while additional experiments are needed to determine the LOD of the protein assay, we believe the limit lies between 5×10^{11} and 1×10^{12} VP/mL. As will be discussed later, the current LOD for the ssDNA assay has been estimated to be $>1 \times 10^{11}$ GC/mL, which will generally place the total protein content within the desired range.

Table 2.2: Estimation of Sample Total Protein Concentration Based on the VP Peak Area under the Curve of the Standard (Known Concentration) and the Area under the Curve of the Sample[a]

Sample No.	Total Concentration (VP/mL)	Protein Area	Predicted Total Concentration (VP/mL)	Prediction Error Rate (%)
Set 1				
1-0	1.07×10^{12}	41.71 ± 1.52	Standard	Standard
1-1	1.07×10^{12}	42.30 ± 3.04	1.08×10^{12}	1
1-2	1.07×10^{12}	43.67 ± 3.40	1.12×10^{12}	5
Set 2				
2-0	1.07×10^{12}	39.60 ± 3.01	Standard	Standard
2-1	1.20×10^{12}	41.16 ± 2.37	1.11×10^{12}	7
2-2	1.33×10^{12}	42.90 ± 2.68	1.12×10^{12}	16
Set 3				
3-0	1.07×10^{12}	31.32 ± 1.68	Standard	Standard
3-1	1.07×10^{12}	31.86 ± 1.08	1.08×10^{12}	2
3-2	1.07×10^{12}	33.52 ± 2.77	1.14×10^{12}	7

[a]The predicted total concentration was estimated using eq 2.3. Moreover, note that each set refers to an independent run in which samples were analyzed in triplicate.

2.4.2 ssDNA Profile Characterization

After confirming that the amount of capsid protein is preserved between full and empty capsids, the next step was to assess the ssDNA content of the capsids (Figure 2.3). To do this, the sample concentration was normalized to 1.07×10^{12} VP/mL and then digested using a standard Proteinase K digestion protocol (refer to Figure 2.1 and Section 2.3.2 Methods). As seen in Figure 2.3, the full reference standard (75% full) exhibits a peak at around 62 s, while the empty standard (96% empty) failed to produce a peak. While it is not surprising that the empty sample did not have a peak, it must be noted that the empty sample should still contain approximately 4.26×10^{10} full capsids, which implies that the LOD is greater than that. Additional experiments were conducted establishing the L.O.D. at approximately 1×10^{11} GC/mL.

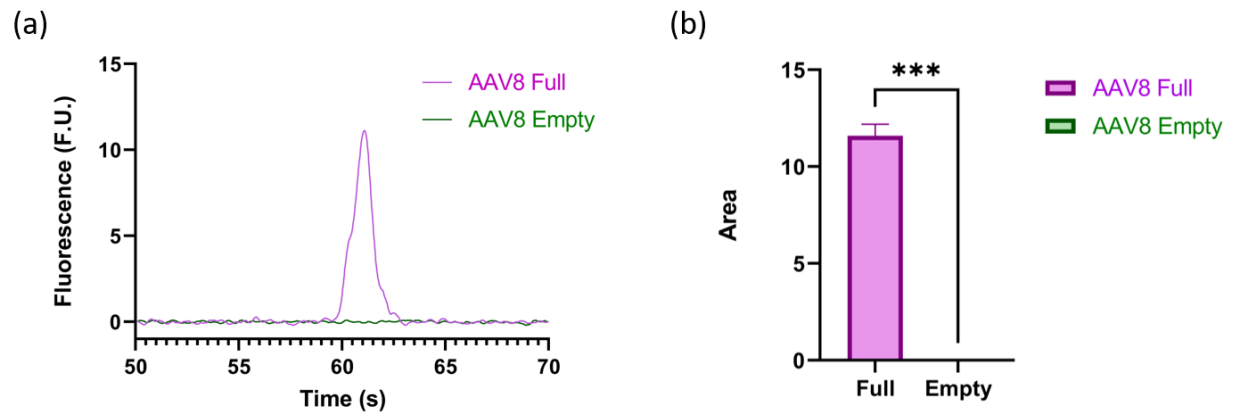

Figure 2.3: Analysis of genomic material from full (purple) and empty (green) AAV8 capsids. (a) Representative electropherogram of full and empty genomic material where the full sample has a peak and the empty sample does not. (b) Summarized area under the curve for the genetic material peak of each sample; note that for all three runs the empty AAV sample did not produce a peak, showing a significant statistical difference between the ssDNA profiles of the two samples.

Like the total protein estimation shown in Table 2.2, the DNA information collected with this assay can be input into eq 2.4 to estimate the number of full particles (or genome copies, GC) in the sample, as seen in Table 2.3. Using this approach, the prediction error rate ranged from 2 to 20%, with an average error rate of 8%. Moreover, it must be noted that this information can be of particular interest for *in vivo* studies as only the full capsids will be carrying the genetic material of interest for gene therapy purposes. However, as previously mentioned, the use of this approach to obtain absolute values rather than relative values normalized to the standard will provide a lower accuracy and is meant to be used as a reference.

Table 2.3: Estimation of Sample Genomic Concentration Based on the ssDNA Area under the Curve of the Standard (Known Concentration) and the Area under the Curve of the Sample[a]

Sample No.	Genomic Concentration (GC/mL)	DNA Area	Predicted Genomic Concentration (GC/mL)	Prediction Error Rate (%)
Set 1				
1-0	7.97×10^{11}	11.32 ± 0.33	Standard	Standard
1-1	5.33×10^{11}	8.02 ± 0.64	5.64×10^{11}	6
1-2	2.66×10^{11}	3.58 ± 0.52	2.52×10^{11}	5
Set 2				
2-0	7.97×10^{11}	10.76 ± 0.32	Standard	Standard
2-1	5.38×10^{11}	7.12 ± 0.70	5.27×10^{11}	2
2-2	2.77×10^{11}	3.64 ± 0.60	2.70×10^{11}	3
Set 3				
3-0	7.97×10^{11}	22.47 ± 2.57	Standard	Standard
3-1	5.33×10^{11}	13.55 ± 2.00	4.80×10^{11}	10
3-2	2.66×10^{11}	6.00 ± 1.61	2.13×10^{11}	20

[a]The predicted genomic concentration was estimated using eq 2.4. Moreover, each set refers to an independent run in which samples were analyzed in triplicate.

2.4.3 Capsid Protein and ssDNA Profiles of Samples with Varying Full Percentages

Next, in order to develop a robust method for estimating the percentage of full capsids in an unknown AAV8 sample, information was collected from samples with varying percentages of full capsids, from 75% to 4% (Figure 2.4). To do this, both samples were normalized to a total protein concentration of 1.07×10^{12} VP/mL, and the full reference standard was mixed with the empty standard at different ratios to achieve the desired percentages of full. As expected, while the protein concentration remains approximately the same with an average total VP area of 28.6–34.16 (Figure 2.4a,b), the ssDNA concentration decreases as the number of full capsids decreases, with average areas going from 11.4 to 0.0 (Figure 2.4c,d).

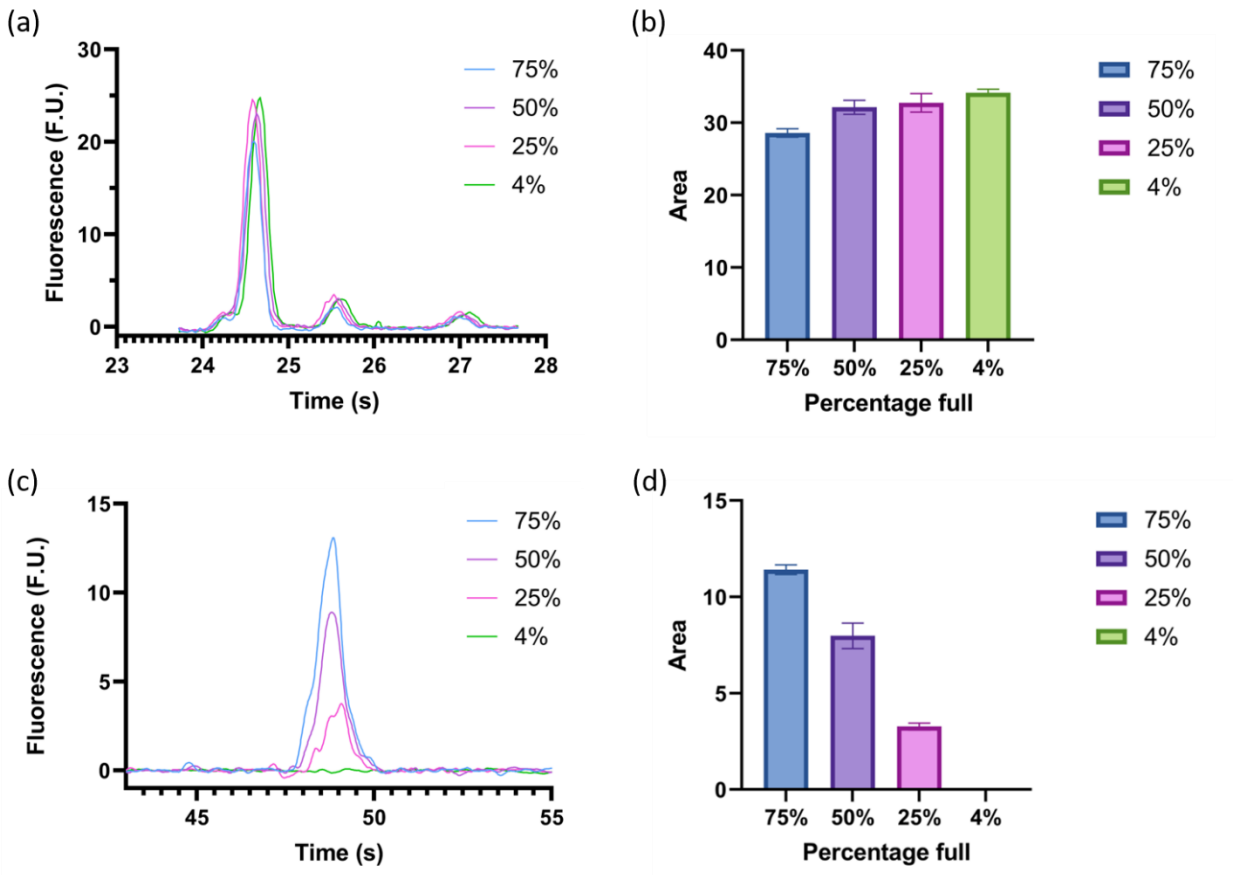

Figure 2.4: Protein and genetic analysis of samples with different percentages of full capsids, including 75% (blue), 50% (purple), 25% (pink), and 4% (green). (a) Electropherogram representative of VP3, VP2, and VP1 capsid proteins from smallest to largest weight (left to right) of both full and empty samples. (b) Summarized integrated area under the curve of the VP protein peaks for the four samples of varying full capsid percentages. (c) Representative electropherogram of the genetic material of the four samples, where the ssDNA peak is observed to decrease as the percentage of full capsids decreases. (d) Summarized area under the curve for the genetic material peak of each sample, for all three runs the empty AAV sample did not produce a peak.

Experimental data was collected on three independent sets of protein and ssDNA assays, analyzed in triplicate. Using the protein and ssDNA information from the full reference standard (75% full) and from the samples, eq 2.6 was used to estimate β, or the predicted percentage of full capsids. The results for each set can be seen in Table 2.4.

Table 2.4: Compiled Protein and ssDNA Data Collected and Analyzed from Three Different Sets of Experiments

Sample No.	Percentage Full (%)	Protein Area	DNA Area	Predicted Percentage Full (%)	Prediction Error (%)
Set 1					
1-0	75	41.71 ± 1.52	11.32 ± 0.33	Standard	Standard
1-1	50	42.30 ± 3.04	8.02 ± 0.64	52	2
1-2	25	43.67 ± 3.40	3.58 ± 0.52	23	2
Set 2					
2-0	75	23.76 ± 2.33	10.76 ± 0.32	Standard	Standard
2-1	45	23.97 ± 2.63	7.12 ± 0.70	48	3
2-2	21	24.01 ± 1.67	3.64 ± 0.60	24	3
Set 3					
3-0	75	22.16 ± 1.12	22.47 ± 2.57	Standard	Standard
3-1	50	22.94 ± 1.15	13.55 ± 2.00	44	6
3-2	25	24.93 ± 2.36	6.00 ± 1.61	19	6

Currently, the average prediction deviation of the model from the actual percentage is ± 4%, ranging from 2 to 6% with a standard deviation of 2%. Alternatively, if the sample concentration is known within an acceptable margin of error, the error and sample analysis time of the system can be significantly decreased by bypassing the protein assay. Assuming an error rate of 10% for the provided concentrations and full percentages of the samples and consequently of our predictions since they are based on these values and the reported error rate of the detection platform, we were able to use the following modified version of eq 2.6 to estimate the percentage of full capsids using the ssDNA area and the provided concentrations

$$R_{concentration} = \frac{c}{c_s} \tag{2.7}$$

$$\beta = \frac{N(f)}{N(f)+N(e)} = \frac{\beta_s R_{DNA}}{R_{concentration}} \tag{2.8}$$

where $R_{\text{concentration}}$ is used to replace R_{p} since, as suggested by the literature and Table 2.3, the sample concentration is related to the protein area.

To assess the specificity of this alternative approach, the actual (reported) percentage of full capsids of each sample was compared to the predictions obtained using eq 2.6, referred to as "Protein", and eq 2.8, referred to as "Concentration", as highlighted in Figure 2.5. Note that while this method bypasses the need for protein analysis, it is still highly dependent upon the ssDNA analysis, as suggested by eq 2.8.

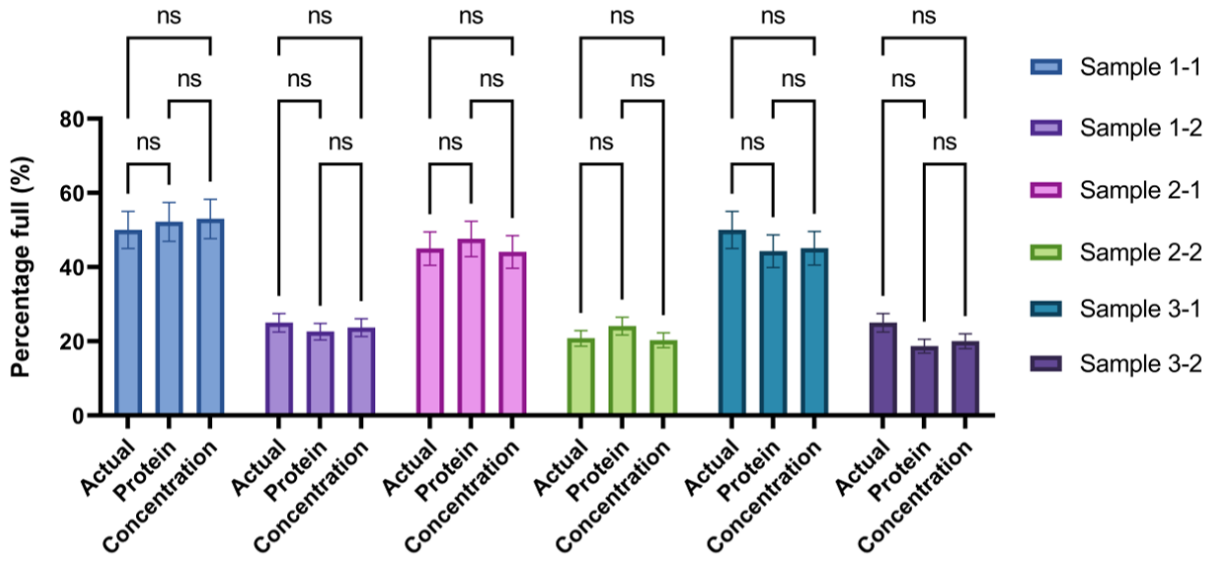

Figure 2.5: Comparison of prediction accuracy between the protein area prediction method based on eq 6 and the concentration prediction method based on eq 8 in comparison to the actual or reported percentage of full capsids of each sample. In the legend on the right, the sample numbers correspond to those described in Tables 2–4. Moreover, all samples include an error bar of 10% (in each direction) of the total value to account for inaccuracies in the reported standards used to prepare both the standard and samples used in this study.

After analyzing the results obtained using both approaches, a decrease from ±4% (protein area prediction method) to ±3% in average prediction deviation was observed when using the concentration prediction method, ranging from 1 to 9%. In other words, when the total sample concentration is provided, both the error and turnaround time can be decreased. Importantly, the decrease in turnaround time will be limited since the longest incubation is needed for the extraction

of ssDNA from the capsid. However, it must be noted that when the known concentrations were used the standard deviation of the predictions increased slightly from 2% to 3%.

2.5 Conclusions

Initially, a single electrophoresis-based microfluidic assay capable of differentiating full from empty capsids was considered in which the difference in their respective isoelectric points (pI) was exploited using a charge variant assay. While both full and empty capsids are assembled in the same fashion, it is believed that the presence of the ssDNA genome in full capsids induces a conformational change that affects its external properties.[90] However, when the full and the empty reference standards were analyzed simultaneously, despite their difference in pI, the samples coeluted even when numerous external factors were used (Table S2.1-2.2 and Figure S2.1) from the Supporting Information). Therefore, the goal of this study shifted to understanding the relationship between the protein and ssDNA profiles of full and empty standards to develop a mathematical model to estimate the percentage of full capsids in an AAV8 sample of unknown concentration and composition.

First, our primary assays were validated by comparing our findings to the literature. As expected, for a given concentration, the protein profiles of full and empty AAV8 did not show a significant difference between them while the ssDNA assay showed a peak representing the ssDNA genome for the full standard and an absence of peak for the empty standard. A major benefit of running these two assays simultaneously is the ability to obtain both the total sample concentration and the total genomic concentration of the sample, as reported in Tables 2.2 and 2.3, in addition to the fraction of full capsids.

Once the individual methods were validated, the relationship between the protein and ssDNA profiles was assessed to develop a mathematical model to characterize the percentage of full

capsids in a given sample. While AUC is still the golden standard when it comes to accuracy, its extensive turnaround time and sample usage leaves room for improvement. The proposed method offers a fast, high-throughput alternative that can be used to quickly iterate through batches of sample without the need for highly specialized equipment that requires significant levels of training. Within 2–3 h, including the denaturing and digestion times, depending on the number of samples being analyzed, the proposed method can predict the percentage of full capsids with an average prediction error of ±4% using a total volume of less than 10 μL per sample (including triplicate analysis). For instance, if 10 samples are being analyzed, it would take an average of 15 min/sample, and if run in triplicate it would take 6 min/sample, which is significantly faster than the current methods (15 min −6 h/sample). Due to the high throughput nature of the proposed method, the latter value would decrease as the number of samples increases, which is particularly relevant for the analysis of samples at different stages of manufacturing (i.e., to assess the effect of each step) and for batch-to-batch analysis. Over the course of these experiments, the LOD of the protein assay was estimated to be between $5 \times 10^{11} – 1 \times 10^{12}$ VP/mL, while the LOD of the ssDNA assay was estimated to be $>1 \times 10^{11}$ GC/mL. However, additional experiments are needed to further narrow down these ranges and determine the limit of quantitation. Moreover, ongoing preliminary experiments with AAV9 (Table S2.3 and Figure S2.3) suggest the method may be compatible with additional serotypes. However, additional experiments are needed to fully assess the cross-serotype compatibility.

2.6 Acknowledgments

This work was in part supported by PerkinElmer's research grant to Brown University. Additionally, we acknowledge James White and others from PerkinElmer for their guidance and

material supplies. The proprietary subjects disclosed herein are patent-pending property of PerkinElmer Health Sciences, Inc.

2.7 Supporting Information

Materials and Methods for Protein Charge Variant Experiments

For the AAV experiments below, the full AAV2 and AAV8 samples are Welgen (Welgen, Inc., Worcester, MA) null controls, at a stock concentration of 5 x 1012 VP/mL, while the empty AAV2 sample is the Vigene Biosciences (Vigene Biosciences, Rockville, MD) empty reference standard, with a stock concentration of 1.27 x 1012 VP/mL but concentrated using the Amicon (MilliporeSigma, Burlington, MA) 50 kDa Filter. The Please note that the exact concentration was not assessed in this section, and that the concentration of sample volume requirements of these experiments are greater than those required by ProteinExpress and PicoRNA. The detection system used in these experiments is the GX Touch II LabChip system (Perkin Elmer, Waltham, MA), and the assay followed was their protein charge variant assay with modifications to the sample labeling protocol (Perkin Elmer). The different pH mixtures were made following the protein charge variant user guide and using the provided pH 5.6 and pH 7.2 buffers.

Table S2.1: Effect of Different pH Conditions on AAV2 Differentiation

Table S2.1: Effect of different pH conditions on full and empty AAV2 differentiation. As seen in the table below, despite the changes in pH, there is no statistically significant difference between the migration time of the full and the empty AAV particles at a pH of 5.6 or 6.5 within our system.

pH	Sample	Migration time (s)
5.6	Full	38.46 ± 0.33
	Empty	38.61 ± 0.33
6.5	Full	57.38 ± 0.59
	Empty	57.36 ± 0.66

Figure S2.1: Electropherograms of AAV2 Full, Empty, and Combined Samples

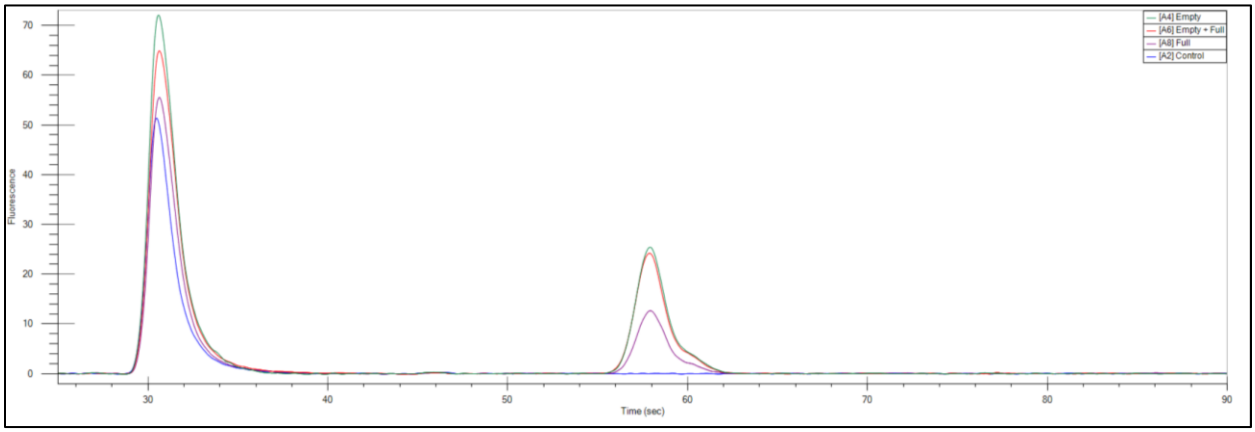

Figure S2.1: Electropherogram of control (blue), full (purple), empty (green), and empty + full (red) AAV2 samples analyzed at a pH of 6.5. The first peak (left to right) is the dye peak, present in the control which lacks any AAV particles, while the second peak is where AAV has been observed to migrate within our system. It must be noted that changes in dye peak magnitude have been observed across the experiments but are not believed to affect the results, and the changes in AAV peak magnitude are a result of differences in concentration between the samples. Moreover, this electropherogram is also a representation of the lack of differentiation or resolution between the empty and full sample peaks, as highlighted by the empty + full sample (red). A similar lack of differentiation was observed across all charge variant experiments regardless of the conditions or compounds used.

Table S2.2: Effect of Different Compounds on AAV2 Differentiation

Table S2.2: Effect of different compounds on AAV2 differentiation. Since changes in pH (Table S2.1 and Fig. S2.1) and in gel-matrix (experiments not included) did not differentiate full from empty AAV particles within our system, the next approach was to induce biochemical changes. Based on literature, antibodies (A1 and A69), digestion compounds (α-chymotrypsin, thermolysin, trypsin), as well as other compounds (PhosBind + streptavidin, urea, SDS) were used to see if they would induce a differential change or bind differentially to AAV in a manner that would result in a difference in mobility, represented by migration time in the table. However, despite the compounds and heat treatments used, we were unable to separate the full and empty peaks from each other. Interestingly, the only compound that appeared to significantly affect the mobility of AAV was α-chymotrypsin, and trypsin appeared to completely digest AAV at the concentration used in these experiments (0.02%) regardless of the incubation time (10-60 min). Please note that different temperatures and incubation times were followed for most compounds, but these were found to be the most representative for each compound.

Compound	Sample	pH	Heat treatment	Migration time (s)
Antibody A1	Full	5.6	65°C x 30 min	39.16 ± 0.02
	Empty			39.12 ± 0.19
Antibody A69	Full	5.6	65°C x 30 min	39.04 ± 0.11
	Empty			39.12 ± 0.19
α-chymotrypsin	Full	5.6	50°C x 30 min	32.34 ± 0.17
	Empty			32.33 ± 0.31
Thermolysin	Full	5.6	50°C x 30 min	38.51 ± 0.34
	Empty			38.58 ± 0.30
Trypsin	Empty	5.6	37°C x 10 min	No peak
	Empty			No peak
PhosBind + Streptavidin	Full	5.6	(RT) 25°C x 20 min	39.00 ± 0.03
	Empty			39.00 ± 0.03
Urea	Full	5.6	(RT) 25°C x 30 min	38.75 ± 0.08
	Empty			38.78 ± 0.07
SDS (under CMC*)	Full	5.6	50°C x 30 min	39.68 ± 0.45
	Empty			39.45 ± 0.36

* CMC = critical micelle concentration

Figure S2.2: Electropherograms of AAV2 and AAV8 Samples

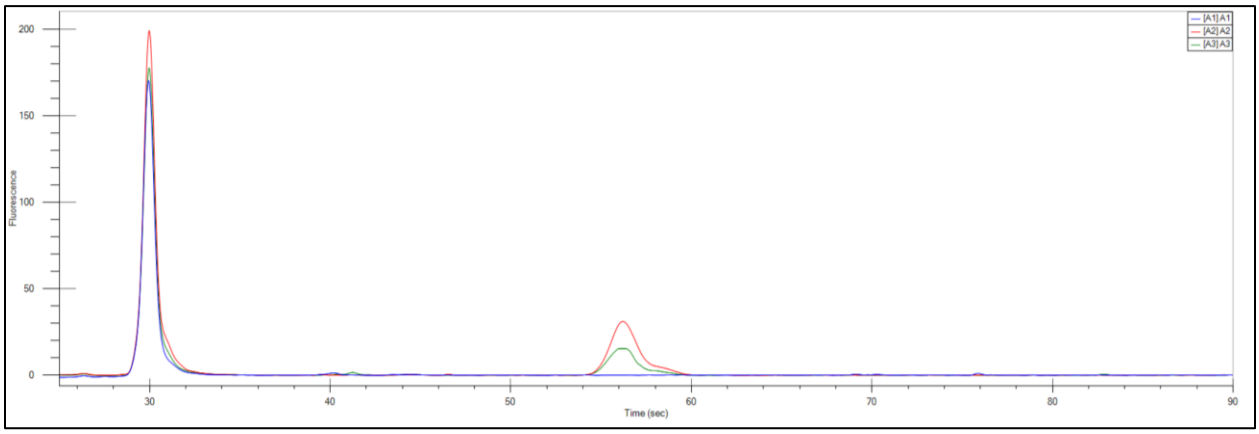

Figure S2.2: Electropherogram of control (blue), AAV2 (red), and AAV8 (green) samples analyzed at a pH of 6.5. The first peak (left to right) is the dye peak, present in all samples including the control which lacks any AAV particles, while the second peak is where AAV has been observed to migrate within our system. It must be noted that changes in dye peak magnitude have been observed across the experiments but are not believed to affect the results, and the changes in AAV peak magnitude are a result of differences in concentration between the samples or to differences in serotype interaction with the dye molecules. Moreover, this electropherogram is also a representation of the similarity of electrophoretic mobility profiles between AAV2 (peak center at 56.20 s) and AAV8 (peak center at 56.07 s). This similarity, which was also observed at different pH values, suggests the AAV2 results above may also be representative of AAV8 behavior. This is further supported by the fact that both samples are believed to contain a mixture of full and empty particles, which do not appear to be differentiated in this electropherogram. However, further studies need to be conducted prior to reaching additional conclusions.

Table S2.3: AAV9 compiled protein and ssDNA data collected and analyzed using mathematical prediction method

Table S2.3: The robustness of the proposed methodology was assessed by using it to predict the percentage of full capsids of a series of AAV9 samples made by combining an AAV9 full and empty reference standard at different ratios (Vigene Biosciences, Rockville, MD). Representative electropherograms for each condition are shown in Figure S2.3. The average prediction deviation for AAV9 was 3%.

Sample #	Percentage Full	Protein Area	DNA Area	Predicted Percentage Full	Prediction Deviation
1-0	82%	11.24 ± 0.23	18.78 ± 1.62	Standard	Standard
1-1	50%	7.41 ± 0.49	7.39 ± 0.77	49%	1%
1-2	25%	6.41 ± 0.32	2.56 ± 0.67*	20%	5%

*This value was estimated from duplicates whereas all others were estimated from triplicates due to the presence of the bubble in the detection channel at the time of analysis.

Figure S2.3: Electropherograms of AAV9 protein and ssDNA assays

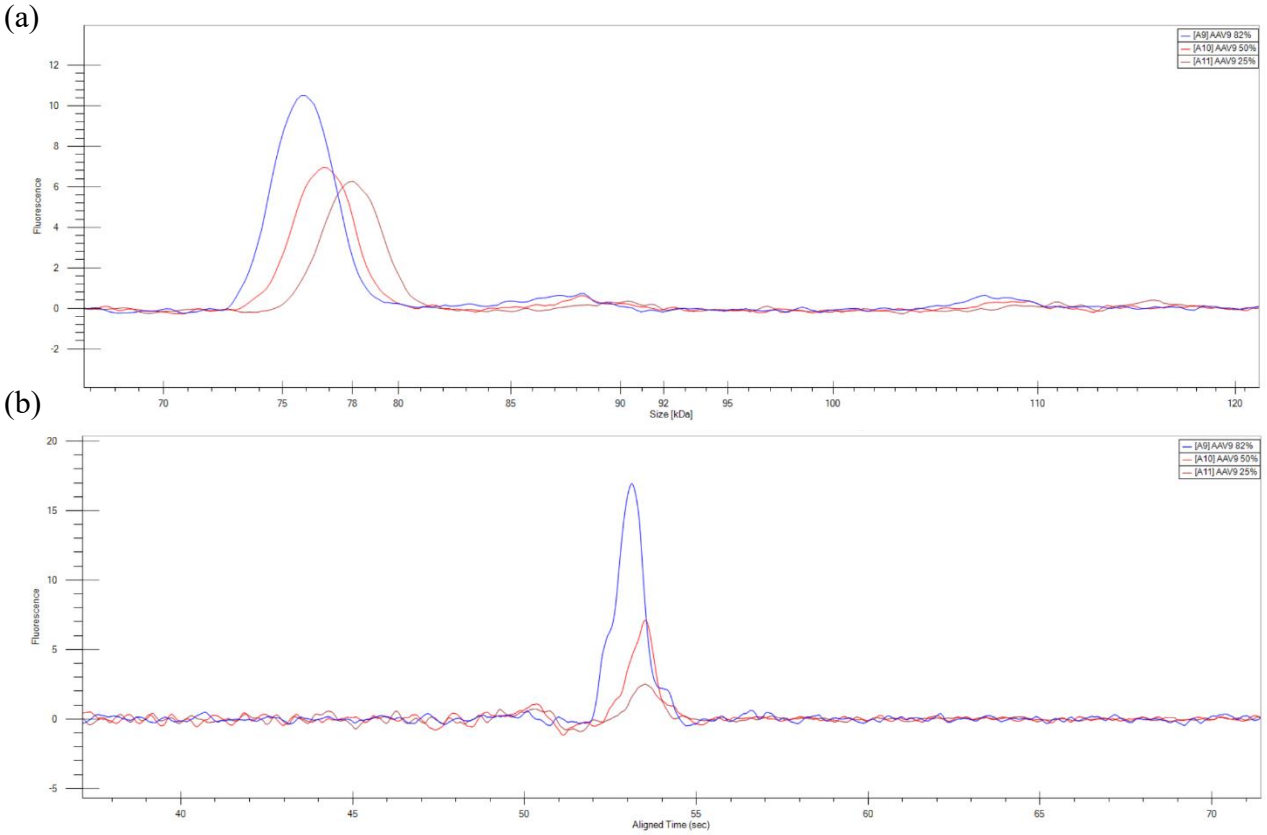

Figure S2.3: Representative AAV9 electropherograms of the (a) protein assay with the 82% full reference standard (blue), a 50% full sample (red) and 25% full sample (burgundy), and of the (b) ssDNA assay with the 82% full reference standard (blue), a 50% full sample (red) and 25% full sample (burgundy).

Chapter 3: Microfluidic AAV Purity Characterization: New Insights into Serotype and Sample Treatment Variability

Chapter 3 has been accepted for publication as an original research article.

Coll De Peña, A.,[1,+] White, J.D.,[2,+] Mehta, D.R.,[2,+] Ben Frej, M.,[2] Tripathi, A.,[1]
Microfluidic AAV Purity Characterization: New Insights into Serotype and Sample Treatment Variability, **ACS Omega** (accepted).

[1]Center for Biomedical Engineering, School of Engineering, Brown University, Providence, RI, USA

[2]Applied Genomics, Revvity, Hopkinton, MA, USA

[+]These authors contributed equally

3.1 Abstract

Despite recent advances in nucleic acid delivery systems with the success of LNP vehicles, adeno-associated virus (AAV) remains the leading platform for targeted gene delivery due to its low immunogenicity to humans, high transduction efficiency, and range of serotypes with varying tropisms. Depending on the therapeutic goals and serotype used, different production conditions may be more amenable, generating an ever-growing need for rapid yet robust analytical techniques to support the high quality manufacturing of AAV. A critical bottleneck exists for assessing full capsids where rapid, high-throughput techniques capable of analyzing a range of serotypes are needed. Here, we present a rapid, high-throughput analytical technique, microfluidic electrophoresis, for the assessment of full capsids compatible with AAV1, AAV2, AAV6, AAV8, and AAV9 without the need for assay modifications or optimizations, and AAV5 with some constraints. The method presented in this study uses a mathematical formulation we developed previously with a reference standard to combine the independently obtained capsid protein and single stranded DNA (ssDNA) profiles to estimate the percentage of full capsids in a sample of unknown concentration. We assessed the ability to use a single serotype (AAV8) as the reference standard regardless of the serotype of the sample being analyzed so long as the melting temperature (Tm) of the capsids is within 12 °C from the Tm of AAV8. Using this method, we are able to characterize samples ± 6.1% with an average analytical turnaround time of < 5 min/sample, using only 10 μL/sample at a concentration of 2.5×10^{12} VG/mL.

3.2 Introduction

Over 30 years after the first gene therapy, there has been a significant increase in the arsenal of therapies and modalities, such as viral (i.e., lentivirus) and nonviral (i.e., lipid nanoparticle), available. Recombinant adeno-associated viruses (rAAV) have become one of the leading platforms for gene therapy. With a diameter of roughly 25 nm[91] and a single-stranded DNA (ssDNA) genome of 4.7 kb,[92] the characteristics that set them apart are their long-term gene expression[93], lack of pathogenicity in humans,[93] and selective tropisms.[93–95] Consequently, AAV characterization remains high on the bioanalytical research agenda, and significant resources have been dedicated to developing methods to assess critical quality attributes (CQAs) to meet the standards and regulations of the FDA. Currently, the most common characterization methods are digital droplet PCR (ddPCR) for the genome titer,[96–98] enzyme-linked immunosorbent assay (ELISA) for the capsid titer,[99–101] and transmission electron microscopy (TEM) for the content ratio in terms of full and empty capsids.[96,102,103] Notwithstanding, analytical ultracentrifugation (AUC) remains the golden standard, particularly in industry, for content ratio assessment due to its increased accuracy and ability to quantify partially filled capsids.[102,104–106] Some other notable techniques include charge-detection mass spectrometry (CDMS),[107] size-exclusion chromatography multi-angle light scattering (SEC-MALS),[108] anion-exchange chromatography (AEC),[81] and mass photometry[87] among others.[106]

While many techniques have been developed and optimized to characterize AAV samples, most require high sample volumes, are low throughput, or have high turnaround times. Therefore, a need remains to support the streamlined analysis of the numerous samples synthesized during the production of AAV and large-scale batch-to-batch analysis. One alternative solution to this need are miniaturized, high-throughput compatible platforms, such as microfluidics. Microfluidics is a

particularly attractive technology due to its precise control of fluids, high-throughput capabilities, and rapid sample processing, making it capable of outperforming traditional technologies.[51] Additionally, the use of high-resolution miniaturized platforms may lower sample and reagent volumes and costs.[6] As a result, depending on their intended use, they can replace or complement many traditional techniques used to characterize AAVs. For instance, this type of platform (micron or submicron) has been used to assess the genomic content,[109] capsid proteins,[109,110] and content ratio of AAV samples.[109,111] However, rather than focusing on the development of a new analytical method, the focus of this study is to increase the fundamental understanding of how different serotypes and sample treatments affect AAV behavior in electrokinetic microfluidic systems and categorize optimized methods based on serotype stability and overall response.

In a previous study, we successfully determined the percentage of full capsids in AAV8 samples of unknown concentration using an AAV8 standard.[109] However, despite the great potential shown by the method, it was limited by (1) using a standard that matched the serotype of the sample, (2) the limited translation to other serotypes due to serotype-specific capsid thermal stability, and (3) the requirement for a time-consuming sample treatment step for nucleic acid extraction. These limitations are important since 10-13 naturally occurring AAV serotypes and over 100 variants have been purified and studied for gene delivery.[93,112,113] The wide range of serotypes and variants makes AAV highly versatile, as they may express different tropisms, or tissue targets, based on their binding receptors (Table 3.1).[93,112,113] To further increase their therapeutic robustness, it has been demonstrated that AAV vectors could be equipped with the desired high-affinity ligands, resulting in increased efficiency targeting the desired cells.[114] Having such an extensive repertoire of AAV vehicles is of great clinical significance as it has been reported that over 90% of humans have been infected with AAV, approximately 50% of which may have neutralizing antibodies

against the virus.[93,112,115,116] As such, serotype or variant selection may significantly impact therapy efficiency based on its ability to evade the immune system. Therefore, it is essential to make a versatile analytical platform that enables the purity analysis of different AAV serotypes for maximal applicability, particularly in a rapid, high-throughput, low-volume format that can support manufacturing optimization and downstream purity assessment.

Table 3.1: Description of the melting temperature, Tm, of the capsids of each serotype,[117–119] the organs they target,[120,121] and the number of clinical trials using that serotype.[122] The capsid Tm was determined elsewhere using differential scanning fluorescence (DSF),[117–119] and the values are reported here as a range combining full and empty capsid values. The number of clinical trials using each serotype was based on queries using the keywords "adeno-associated virus" and "AAVX", where "X" represents the serotype number on the clinicaltrials.gov website. The search using "adeno-associated virus" yielded 180 trials, and other studies have reported as many as 331 trials,[123] whereas the sum of the clinical trials included below is 162. This discrepancy is due to the lack of specificity in the description or use of different variants in the trials.

Serotype	Capsid Tm (ºC)[117–119]	Organs Targeted[120,121]	Number of Clinical Trials[122]
AAV1	82-84.5	Eye, muscle, CNS, heart	12
AAV2	66.33-71.6	Eye, brain, lung, liver, muscle, joint, CNS, kidney	59
AAV3	67.3-71.7		0
AAV4	74.0-75.0	Eye, lung, CNS	0
AAV5	88.07-90.5	Eye, lung, liver, CNS	19
AAV6	77.31-78.5	Lung, muscle	2
AAV7	76.5-76.7	Liver, muscle	0
AAV8	70.59-73	Eye, liver, muscle, CNS, heart, pancreas	30
AAV9	76.17-77.0	Lung, liver, muscle, CNS, heart	40

Our initial method was developed and validated using AAV8,[124] and prior to this study, further assay optimizations were conducted using AAV8 as well; therefore, it was decided to retain AAV8 as the reference standard in this study due to the extent of the prior body of work and the reference standard serotype comparison conducted in this study. Based on the body of scientific literature available and ongoing clinical trials (Table 3.1) for each serotype, it was determined that AAV2, 8, and 9 were the most commonly used in AAV research. After investigating the differences and similarities across the different serotypes, we found that as long as the capsid melting temperature of the serotypes (Table 3.1) was within 12 °C from AAV8, they responded similarly to the chemical, biochemical, and physical (heating) treatment used to characterize the samples in this study. This range should include AAV1, 2, 3, 4, 6, 7, and 9 (Table 3.1). Based on the organs targeted and the number of clinical trials, we validated the method with serotypes 1, 2, 6, 8, and 9, which cover the expected operational temperature range. We also attempted to analyze AAV5 due to its prevalence and number of clinical trials. However, the results were beyond our accepted threshold for prediction accuracy due to the significantly higher capsid Tm, which prevents its proper denaturing and digestion using the current protocol. Notwithstanding, based on its clinical relevance, we evaluated the performance of the assay to predict the percentage of full capsids in AAV5 samples and highlighted the current limitations and a potential solution to overcome them.

Upon narrowing down the functional melting temperature (Tm) range for the assay, we evaluated how different serotypes (1, 2, 6, 8, and 9) performed as the reference standard for the estimation of full and empty capsids. Despite the ranging degrees of success, there was no statistical difference between their efficiency, with AAV1 and AAV9 showing the highest variability. Ultimately, AAV8 was selected as its capsid Tm is roughly in the middle of the most commonly used serotypes (AAV2, 8, and 9). By reducing the time required to extract the nucleic acid from

the capsid and optimizing the nucleic acid and capsid protein assays, we were able to develop a more robust method that enables the rapid characterization of an unknown AAV sample of serotypes 1, 2, 6, 8 and 9 (Figure 3.1) using low volumes (10 µL) within our desired threshold, and AAV5 with a larger error. In addition, the insights provided by the study regarding the relationship between sample Tm and treatment response can provide valuable information to expand the method into a platform method, including proprietary AAV variants for development laboratories. Being able to assess the purity of AAV samples rapidly serves a 2-fold purpose: upstream, it can help optimize the AAV manufacturing conditions (postpurification), while downstream, it can help characterize the final product.

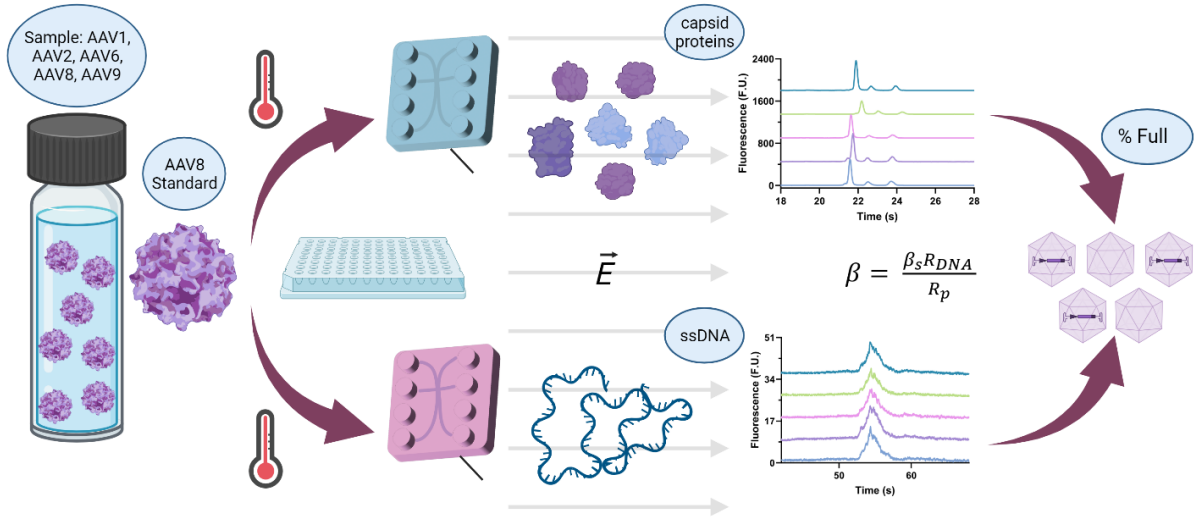

Figure 3.1: Workflow for the microfluidic electrophoresis characterization of AAV1, AAV2, AAV6, AAV8, and AAV9 was calculated by analyzing the capsid and genomic composition separately and integrating the information using a reference standard and a mathematical formulation to estimate the percentage of full capsids in the sample.

3.3 Materials and Methods

3.3.1 Materials

AAV (1, 2, 5, 6, 8, and 9) reference standards were purchased/obtained from Revvity (Waltham, MA) at a concentration of 1×10^{13} GC/mL and were diluted using PBS + 0.001% Pluronic to 2.5

$\times 10^{12}$ GC/mL. Different serotypes either came or were diluted with empty capsids to achieve different percentages of full, which will be specified throughout the study. The LabChip GXII Touch platform, custom AAV protein microchip and chip reagents, and custom AAV ssDNA microchip and chip reagents were obtained from Revvity. The UV/vis percentage of full capsids was estimated by Revvity and provided as a single value, which is why there are no error bars associated with those values.

3.3.2 Methods

The AAV characterization experiments discussed in this study were conducted on the analytical platform LabChip GXII Touch (Revvity) for optics, robotic motion, pressure, and electrokinetic control. Two separate assays were developed or optimized for the purity assessment of the AAV samples: a capsid protein and an ssDNA assay. The samples were treated, heated on a 96- or 384-well plate for both assays and transferred onto the LabChip platform. For each assay, an assay-specific (AAV protein or AAV ssDNA) microfluidic chip was loaded with a gel or a gel-dye and a marker for calibration purposes. The chip was placed on the LabChip platform. Once the well plate and chip are loaded onto the platform, the system moves the desired sample well under a metal sipper attached to the microfluidic chip. By vacuum pressure, 20 nL of each sample was loaded onto the chip individually for analysis. Additional information can be found under Methods for Capsid Protein and ssDNA Characterization in the Supporting Information. The statistical analysis for this study was conducted using GraphPad Prism 9, and the figures were made using GraphPad and/or BioRender.com.

3.3.3 Theory

In a previous study, we proposed a mathematical formulation that enabled us to estimate the percentage of full capsids in a sample of unknown concentration and percentage of full capsids

using a microfluidic electrophoresis method that resulted from the combination of a protein and ssDNA assay that we developed for AAV purity assessment.[124] To achieve this, a key innovation was integrating a reference standard of the known percentage of full capsids and concentration into the workflow to normalize the raw fluorescent data obtained from the electropherograms. For this, we described the total number of particles in a sample, N, and standard, N_s, in terms of the full, f, and empty, e, particles present in the sample[124]

$$N = N(f) + N(e) \tag{3.1}$$

$$N_s = N_s(f) + N_s(e) \tag{3.2}$$

Full and empty capsids are composed of the same protein subunits (VP1-3); therefore, regardless of the genetic content, each capsid should contain the same amount of protein α. Therefore, the relation between the concentration, c, of proteins in the samples and the standard can be represented by[124]

$$R_p = \frac{c(protein)}{c_s(protein)} = \frac{\alpha N}{\alpha N_s} = \frac{N}{N_s}, \tag{3.3}$$

while that of the ssDNA insert present in the sample and standard can be expressed as[124]

$$R_{DNA} = \frac{c(ssDNA)}{c_s(ssDNA)} = \frac{N(f)}{N_s(f)} \tag{3.4}$$

We then defined the known percentage of full capsids in the standard, β_s, as[124]

$$\beta_s = \frac{N_s(f)}{N_s(f)+N_s(e)}, \tag{3.5}$$

which can be used to estimate the percentage of full capsids in the sample through the relation[124]

$$\beta = \frac{N(f)}{N(f)+N(e)} = \beta_s \frac{c(ssDNA)}{c_s(ssDNA)} \div \frac{c(protein)}{c_s(protein)} = \frac{\beta_s R_{DNA}}{R_p} \tag{3.6}$$

By independently analyzing both the protein and ssDNA profiles of the unknown AAV sample(s) in tandem with a reference standard, we can mitigate the error introduced by the use of two independent assays by helping normalize the concentrations yielded by each assay to the standard, allowing us to combine the assays. Using this approach, we were able to estimate the percentage of full capsids in an AAV8 sample using an AAV8 standard ± 4%.[124] Since the basis for this mathematical formulation is independent of serotype and showed a high degree of accuracy in the past, it will also be used to estimate the percentage of full capsids in this study, but the optimal standard serotype will be explored.

3.4 Results and Discussion

3.4.1 Protein Analysis Optimization

In our mathematical formulation, the protein signal (corrected area) is used to determine R_p. This parameter can be obtained using both protein assays described in the methods: the Protein Express and AAV Pico Protein assays. While there are some differences in the sample treatment and chip loading between these two assays, the critical difference that we believe is responsible for the difference in response of the two is the use of a noncovalent dye (Revvity) in the Protein Express assay and a covalent dye (Revvity) in the AAV Pico Protein assay. Prior to this study, we optimized both assays (not shown in this study) by using an AAV8 sample. Using these protocols, we analyzed AAV1, 2, 6, 8, and 9 samples using the noncovalent (Protein Express) and the covalent (AAV Pico Protein) assays (Figure 3.2).

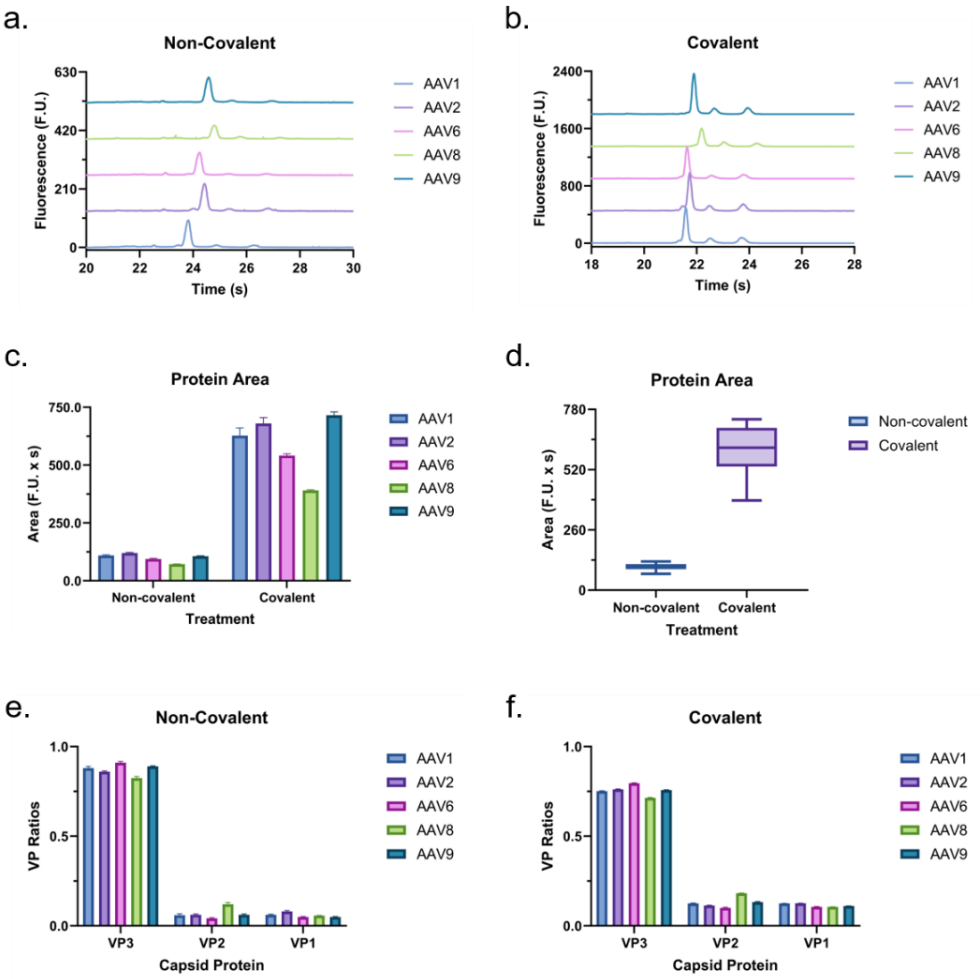

Figure 3.2: Electrophoretic characterization of the capsid proteins of AAV1, 2, 6, 8, and 9. Electropherograms of the different serotypes using (a) the noncovalent dye with an offset of 130 F.U. and (b) the covalent dye with an offset of 450 F.U. The peak areas for all three capsid proteins are highlighted (c) by serotype and (d) by dye type. Comparing the covalent dye response to the noncovalent, we see a 6-fold increase in peak area. Note that all samples were normalized to 2.5×10^{12} VG/mL at different percentages of full and are expected to have different capsid concentrations. However, while the concentrations vary across serotypes, they were used at the same concentration in both assays, enabling a direct comparison between the two. Predicted VP ratios using (e) the noncovalent and (f) the covalent assay. The VP ratios were estimated by dividing each peak area by the protein MW (Table S3.1) and normalizing the values.

When the electropherograms yielded by each assay (Figure 3.2) are compared, we observe over a 6-fold increase in peak areas using the covalent dye from the non-covalent dye, which leads to a greater definition of the viral proteins (VP), in particular of VP1 and VP2. Based on these results, the limit of detection (LOD), as defined by the minimum concentration at which we can detect all

observable peaks for the capsid proteins, is 9.9×10^{11} VP/mL for the non-covalent dye and 6.8×10^{10} VP/mL for the covalent dye, a 14.5-fold difference. Despite not assessing the molecular interaction of the dyes directly, we believe this difference in response stems from the mechanism through which each dye interacts with the capsid proteins. Whereas the non-covalent dye will bind to the proteins through some unspecific charge attraction, the covalent dye has a more specific target with higher affinity, leading to an increase in binding sites and resulting overall labeling efficiency. Interestingly, although the peak areas vary due to differences in sample concentration, we see the same behavior in both assays where AAV2 and AAV9 appear to have the highest capsid concentration and AAV8 the lowest, suggesting that the two assays are in agreement. This response also suggests that while the labeling efficiency may change, the difference in affinity between the two dyes is unlikely to be serotype-biased.

While the primary motivation behind the analysis of the capsids is to determine R_p, the analysis of the denatured capsids also provides great insight into the VP ratios (Figure 3.2e-f). While the exact VP ratio may vary, VP stoichiometry is expected to be 10:1:1 for VP3/VP2/VP1.[125] However, Snijder et al. demonstrated that depending on the concentration at which the VPs are present, capsid symmetry will not be followed, and the assembly will be stochastic.[126] In addition, VP1 contains a domain essential for endosomal escape and, consequently, virus infection.[127,128] Therefore, a decreased ratio of VP1 to total capsid proteins can reduce transduction efficiency.[129] As such, VP ratios are a critical-to-quality metric, making the ability to monitor them of great therapeutic relevance since changes to AAV capsid protein stoichiometry, in particular VP1 amount, can lead to a loss of potency.[127]

In this study, the ratios were estimated to be 10.00:0.80:0.68 and 10.00:1.72:1.51 for the noncovalent and covalent assays, respectively (Figure 3.2e-f). These values were estimated by

normalizing the area of each peak, which is relative to the concentration of each capsid protein and to its molecular weight, to enable a comparison of the relative protein molecules in the sample. While there is a difference between the two, because of the stochastic nature of the ratios, without an orthogonal method to assess the actual VP ratios of the sample, it is difficult to determine which provides the most accurate ratios. However, because of the repeatability of the results, the estimated ratios have great value for batch-to-batch reproducibility. In addition, a comparison of the peak profiles associated with each VP suggests that the covalent method shows a greater ability to observe changes in peak area while producing the most consistent and narrow peaks (Figure S3.1). Another VP property that can be monitored using this method is the size. Using the internal LabChip system protein ladders, we were able to estimate the size of the different capsid proteins, with an average % CV of 1.1 and 0.4% for the noncovalent and covalent methods, respectively (Table S3.1). Interestingly, while the two assays are in close agreement, both seem to overpredict the size of the proteins when compared to the expected sizes of VP3 (59-61 kDa), VP2 (64-67 kDa) and VP1 (79-82 kDa).[64] Due to the magnitude of the difference, possible explanations for this difference are probably a combination of post-translational modifications and protein conformations, which may affect their ability to migrate through the gel matrix of the microfluidic chip.[92] Notwithstanding, despite the differences in expected and observed VP size, both assays yielded highly repeatable sizing values, with the covalent assay outperforming the noncovalent assay based on a slight difference in coefficient of variation.

Based on the findings presented in this section, the covalent method appears to yield the highest sensitivity, which is essential for detecting subtle yet important changes across protein profiles, and the lowest coefficient of variation, which plays a crucial role in batch-to-batch comparisons. The ability to consistently detect subtle changes is also highly relevant for production optimization,

including cell line selection and growth condition optimization. Therefore, the remainder of the study will discuss only the results obtained with the covalent assay.

3.4.2 ssDNA Optimization

In our previous study, we used proteinase K to digest the capsid for ssDNA release prior to analysis.[124] Despite effectively releasing the genomic content of the AAV capsids in a reproducible manner, the treatment requires ~1.5 h of treatment, which takes a significant toll on the turnaround of the method. Therefore, we optimized an assay that requires only a short (~10 min) heat treatment without the use of digesting enzymes. In Figure 3.3, we compare the performance of the two using AAV1, 2, 6, 8, and 9 at a concentration of 2.5×10^{12} VG/mL. Since the genomic concentrations are normalized across all serotypes, we expect all samples to yield equal peak areas here. When we compare the two sample treatments, the heat treatment samples produce slightly higher and sharper peaks, which may be easier to detect at low concentrations. However, the average areas were almost identical. Since the areas produced by each approach are nearly identical in magnitude, so is their LOD of 1.6×10^{11} GC/mL. In contrast, the areas produced by the proteinase K treatment have less variation across the different serotypes. Both sample treatments yielded nearly identical migration times for all serotypes. Overall, the higher variation in peak area when using the heat treatment was outweighed by the benefits of the peak shape, sensitivity, and reduced incubation time. Therefore, the results in the remainder of the study will include only heat-extracted ssDNA data. In addition, while not exploited here, a key benefit of the proposed ssDNA analysis method is that the electrophoretic separation of the genomic content within the microfluidic chip enables genomic-fragment-based analysis of the contents of the capsid so long as the concentrations are within our limit of detection.

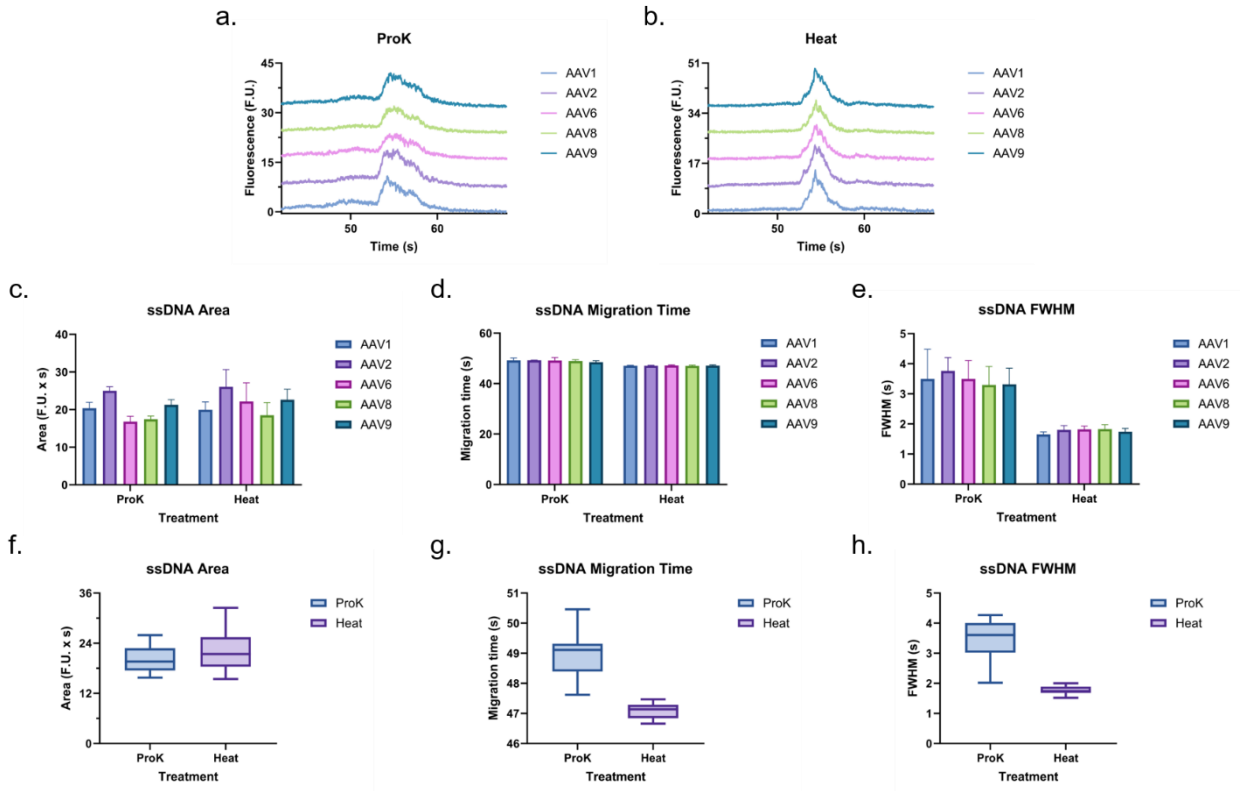

Figure 3.3: AAV ssDNA electropherogram profiles post-treatment with (a) proteinase K (proK, electropherograms offset by 8 F.U.) and (b) heat (electropherograms offset by 10 F.U.). Comparison of peak (c) area, (d) migration time, and (e) fwhm using each treatment separated by serotype. Average peak (f) area, (g) migration time, and (h) fwhm for each treatment.

3.4.3 Determining Optimal Standard and Assay

Once we determined the optimal capsid protein assay and ssDNA extraction method, we used eq 3.6 to assess the performance of five different AAV serotypes (1, 2, 6, 8, and 9) as the reference standard used to estimate the percentage of full capsids in AAV samples of different serotypes. For this, we analyzed the different serotypes simultaneously and used the protein (Figure 3.2) and ssDNA (Figure 3.3) data of each to estimate the percentage of full capsids in the other serotypes (Figure 3.4a). Using the expected percentage of full capsids in each sample, we calculated the average prediction error for each serotype as a standard (Figure 3.4b). While there was no statistically significant difference across the performance of each serotype based on their

prediction deviation, AAV2, AAV6, and AAV8 yielded a narrower distribution. Interestingly, the range of capsid Tm covered by the serotypes used in this section spans from 66.33 to 84.5 °C, whereas AAV2, 6, and 8 have capsid Tm values of 66.33 to 71.6 °C, 77.31 to 78.5 °C and 70.59 to 73 °C, respectively (Table 3.1). In contrast, AAV1 and AAV9 have capsid Tm values of 82 to 84.5 and 76.17 to 77.0 °C, respectively. The results over this temperature range suggest that while temperature appears to play a critical role in the ability for a serotype to be used as a cross-serotype reference standard, the difference in prediction error between AAV6 and AAV9 despite their similar capsid Tm indicates other biochemical features may play a role. Notwithstanding, given its low prediction error, relevant Tm (close to the middle of the range evaluated), and clinical significance, we decided to use AAV8 as the reference standard.

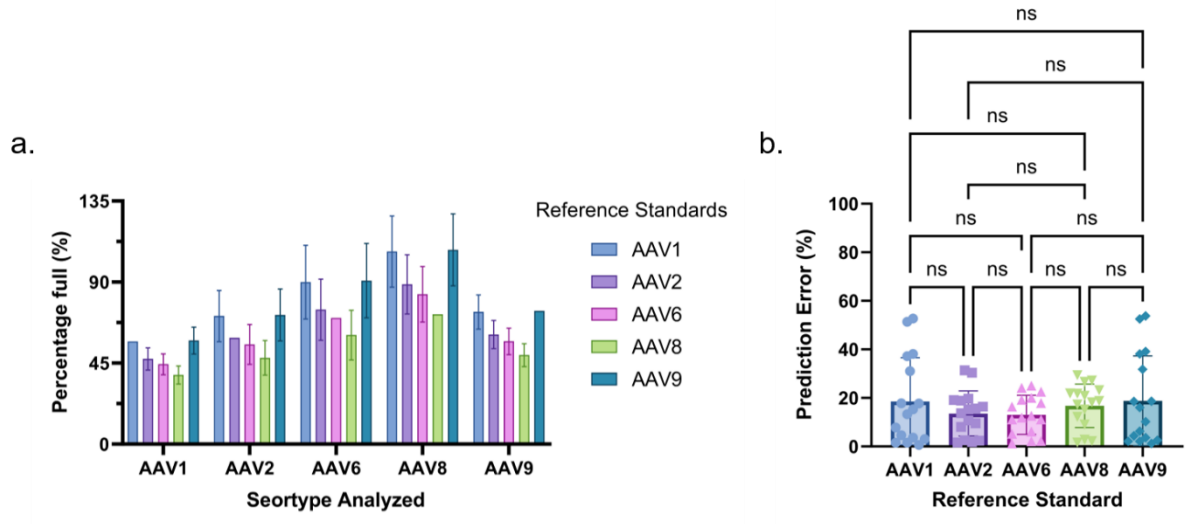

Figure 3.4: Comparison of the predictive performance of each serotype as a reference standard to estimate the percentage of full capsids in the other samples of different serotypes using eq 3.6. (a) Predictions separated by serotype being analyzed, where each color represents the serotype used as the reference standard for those estimations. Note that serotypes were not used to predict themselves, and the expected value was used instead (i.e., the percentage of full capsids in the sample using an AAV1 reference standard is the expected percentage of full capsids in the sample). (b) Summarized prediction error of each serotype as a reference standard.

3.4.4 Application: Assay Performance

To test the performance of the proposed method, we decided to look at a range of serotypes that covered the operational capsid Tm range of the assay and with varying percentages of full capsids. To this end, we analyzed two or three samples of varying percentages of full capsids of AAV1, 2, 6, and 8, covering a range from 66.33 to 84.5 °C (Table 3.1). Based on the earlier findings presented in this study, the final assay uses AAV8 as the reference standard, the covalent dye for protein analysis, the heat treatment for the ssDNA analysis, and the mathematical formulation to make the predictions. In addition, to prevent bias when analyzing AAV8 due to the reference standard selection, independently generated AAV8 preparations were used as the reference standard and as the sample being analyzed. Following the proposed method, the average prediction error was ± 6.1% with a CV of 15.3% (Figure 3.5a, Table S3.2), which is within our ≤ 20% cutoff value.

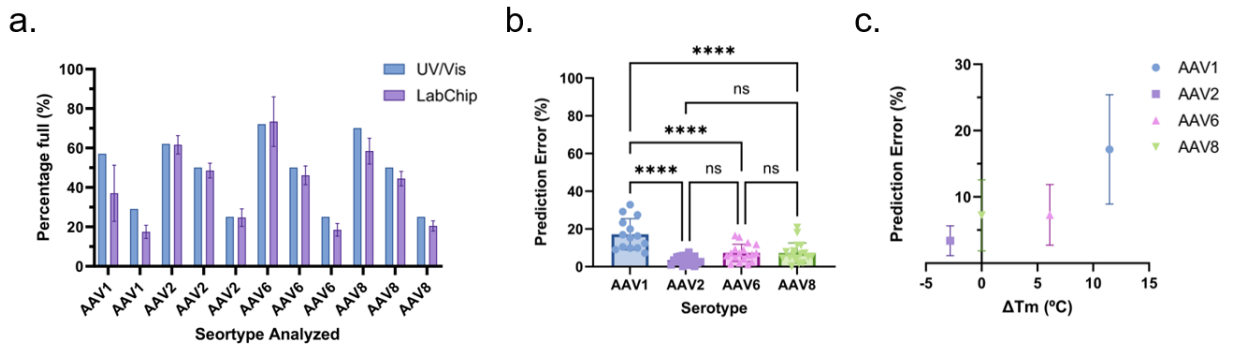

Figure 3.5: (a) Comparison of the proposed method (LabChip) to UV/vis (Stunner). Comparison of prediction errors relative to each (b) serotype (AAV1, 2, 6, and 8) and (c) ΔTm to AAV8 reference standard. For the ΔTm estimations, an average of the ranges provided in Table 3.1 was used for each serotype. No error bars are present in the UV/vis data as this was provided as an average by the provider.

While the overall prediction error and CV were within the desired ranges, when we break the errors down by serotype (Figure 3.5b), we see that the AAV1 predictions underperform compared to the other three serotypes tested in this section. We see an interesting trend when we take it a step further and look at the capsid ΔTm between the reference standard and the samples.

When the ΔTm increases and is positive, the prediction error increases significantly (Figure 3.5c). On the other hand, while the data set is too limited in that direction to draw a meaningful conclusion, it appears that when the ΔTm is negative, the prediction accuracy is less affected by the difference in ΔTm. We also performed an additional comparison between the LabChip and the UV/Vis predictions (Figure S3.2), where most predictions fell within ±10% of each other with the exception of the AAV1 estimations. These findings further support that Tm, which should not affect the UV/Vis-based estimation, may be responsible for the larger errors associated with some serotypes using the LabChip technique.

3.4.5 Application: AAV5

Lastly, while acknowledging that AAV5 (ΔTm = -17.5 °C) is outside the optimal operational range (ΔTm = ± 12 °C) of the proposed assay, given its clinical relevance, we decided to assess the performance and limitations of this method to analyze AAV5. As expected, when the protein (Figure 3.6a) and ssDNA (Figure 3.6b) of three AAV5 dilutions were analyzed, they both showed a strong linear relation between the peak area and concentration. Moreover, when the protein and ssDNA data for these samples was used to predict the percentage of full capsids in the samples, as highlighted in Figure 3.6c, the three samples yielded highly consistent results of 71.0 ± 5.4, 70.3 ± 1.2 and 66.4 ± 3.8% for AAV5-1, AAV5-2, and AAV5-3, respectively. When the predictions for the three samples are combined, the average prediction is 69.1 ± 4.1%. Still, despite the reproducible results, when the predictions are compared to the UV/vis estimation, which was 93%, there is a difference of 23.9%. However, since the results showed to be repeatable and there is a strong linear relation between sample concentration and peak area for AAV5, upon further analysis, a case could be made for using a correction factor in order to increase the accuracy in the estimation of full capsids in AAV5 samples.

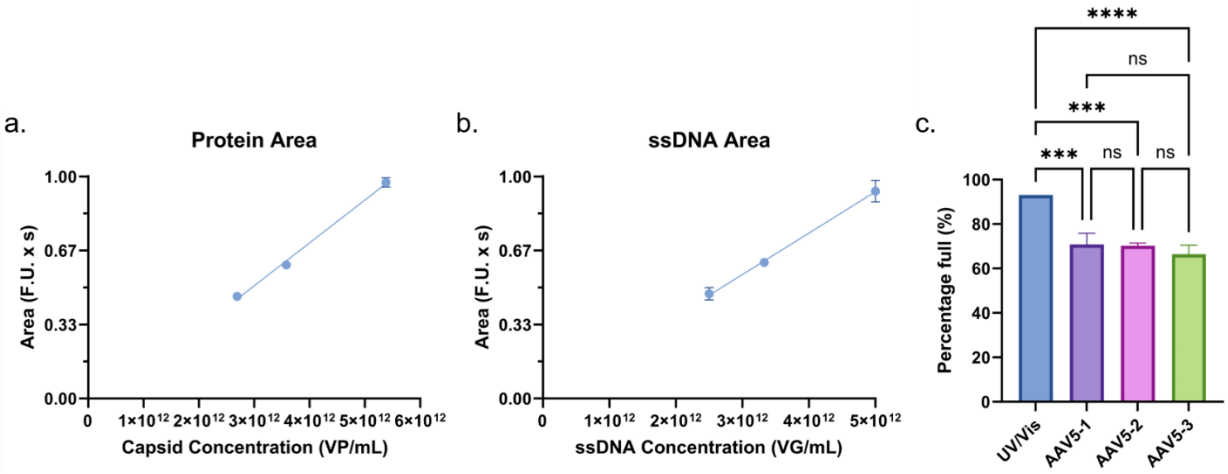

Figure 3.6: (a) Protein and (b) ssDNA peak area over concentration for three dilutions of AAV5 at 93% full per UV/vis measurement, where the protein response has an R^2 of 0.99 and the ssDNA response has an R^2 of 0.98 (c) Predicted percentage of full capsids in the three AAV5 dilutions at 93% full (AAV5-1 is 1:4, AAV5-2 is 1:3, and AAV5-3 is 1:4 of an initial stock of 1×10^{13} VG/mL) using an AAV8 standard. No error bars are present in the UV/vis data as this was provided as an average by the provider.

3.5 Conclusions

In this study, we improved the sensitivity of the protein analysis 12.5-fold to an LOD of 3.4×10^{11} VP/mL by using a covalent-fluorescent dye labeling strategy. Mechanistically, the noncovalent method should be more general and provided VP ratios closer to the expected 10:1:1 ratio but had a larger LOD when compared to the covalent method. The AAV Pico Protein assay yielded a more sequence-dependent ratio but has higher sensitivity and similar repeatability to the noncovalent method. In our view, the covalent method may be developed along with specifications to monitor AAV capsid production. In contrast, the sensitivity of the ssDNA assay did not change with an LOD of 1.6×10^{11} GC/mL. The sample preparation time decreased 8-fold to 10 min, bypassing the need for enzymes. Once the methods to collect the protein and ssDNA information were determined, different serotypes were evaluated for their potential as reference standards. Ultimately, based on the capsid Tm, percentage full prediction error, and clinical relevance, AAV8 was selected as the default reference standard.

The performance of the methods integrated through the reference standard and mathematical formulation was then tested against samples of AAV1, 2, 6, and 8 at different percentages of full capsids, which yielded an average error of ± 6.1% with a CV of 15.3% within a △Tm range of ± 12 °C. The analysis of AAV5 was also attempted, but the difference in capsid Tm led to an average prediction error of ±23.9%, which was above our desired threshold of ±20%. Interestingly, the only instance where the % CV, but not the average error, was above the desired threshold of 20% was with AAV1. Given that AAV1 has the greatest capsid △Tm relative to AAV8, we explored the relationship between the capsid △Tm and prediction error. Not surprisingly, the greater the △Tm, the greater the error. While we have limited data in the negative direction, the data also appear to indicate that negative △Tm may be less impacted by differences in capsid Tm. However, based on the difference in the prediction accuracy of AAV6 and AAV9, which have very similar capsid Tm, additional serotype-specific properties may also play a role.

Overall, this improved method enables the rapid characterization (< 5 min/sample) of the percentage of full capsids in a wide range of serotypes. While traditionally, the combination of two independent methods to determine the genome to capsid ratio may lead to a compounded error, the mathematical formulation we developed uses the reference standard to combine the assays in a way that mitigates this error. In contrast, analyzing the protein and ssDNA content separately enables us to look closer at VP ratio and ssDNA purity and integrity while only requiring 10 µL of sample.

The cross-serotype functionality and ability to address multiple critical quality attributes (CQAs) simultaneously are of great relevance as there are at least 331 clinical trials that include AAV, and a great analytical bottleneck in their development is the lack of reliable, high throughput

characterization methods. An important example of this bottleneck is the lack of robust high throughput analytical tools for assessing the percentage of full capsids in an AAV sample, with AUC, a non-scalable technique, as the current gold standard. Lastly, having an integrated platform that enables the assessment of numerous CQAs, such as total protein concentration, VP ratio, ssDNA concentration and integrity, and percentage of full capsids, can decrease the method transfer burden from analytical development to quality assurance and quality control.

3.6 Acknowledgements

This work was funded by Revvity. A.T. is a paid scientific advisor/consultant and lecturer from Revvity. The authors thank Sirion Biotech (Revvity) for the samples and the UV/vis data.

3.7 Conflict of Interest

The authors declare no competing financial interest.

3.8 Supporting Information

Methods for Capsid Protein and ssDNA Characterization

For the characterization of the capsid proteins, two different protein assay protocols, whose key difference was the use of a fluorescent non-covalent dye (Protein Express, Revvity) or a covalent dye (AAV Pico Protein, Revvity), were followed to investigate how different fluorophore types interact with the capsid proteins. The non-covalent dye protocol, as previously described in Coll De Peña et al.,23 required the mixing of 5 µL of the sample with 7 µL of denaturing reducing buffer with 0.03 M DTT, heated at 100°C for 5 min, and then diluted with 32 µL of nuclease-free water prior to analysis. Since the labeling in this protocol is dynamic and takes place on the chip, the chip is loaded with a gel-dye mixture and a lower marker (Revvity) and analyzed using the Protein Express assay script. In the case of the covalent dye, 4 µL of AAV sample was mixed with 1 µL of 5x Labeling Buffer with 0.17 M DTT (Revvity) and denatured at 75°C for 5 min. The denatures samples were mixed with 5 µL of 20 µM covalent dye (Revvity), shake mixed for 1 min, spun down for 1 min at 1,000 g, and incubated at 35°C for 15 min in the dark. Lastly, 5 µL of the labeled sample were mixed in a 96- or 384-well plate with 5 µL of Stop Solution (Revvity), shake mixed for 1 min, spun down, and transferred to the LabChip platform. The microfluidic chip was loaded with AAV Pico Protein Express gel matrix and lower marker (Revvity) and transferred to the LabChip platform for analysis using the AAV Pico Protein Express assay script.

In contrast, to assess the nucleic acid content of the AAV samples, the capsid was digested via a urea and proteinase K treatment with heat, as described in our previous study,23 or only heat to extract the NA. In the case of the urea and proteinase K treatment, 10 µL of proteinase K was diluted with 90 µL of 2M urea, and 5 µL of that mixture was added to 5 µL of AAV sample. The samples were heated at 55°C for 60 min, following the deactivation of proteinase K at 95°C for 20

min. The samples are then transferred onto a 384-well plate and are ready for analysis. For the heat treatment samples, 5-10 µL of AAV are heated at 94°C for 10 min, and upon transfer to a 384-well plate, the samples are ready for analysis. For both sample treatment protocols, the microfluidic chip is loaded with AAV DNA gel-dye and a lower marker (Revvity) and analyzed using the AAV DNA assay script.

Figure S3.1: Comparison of VP Properties Across the Two Methods

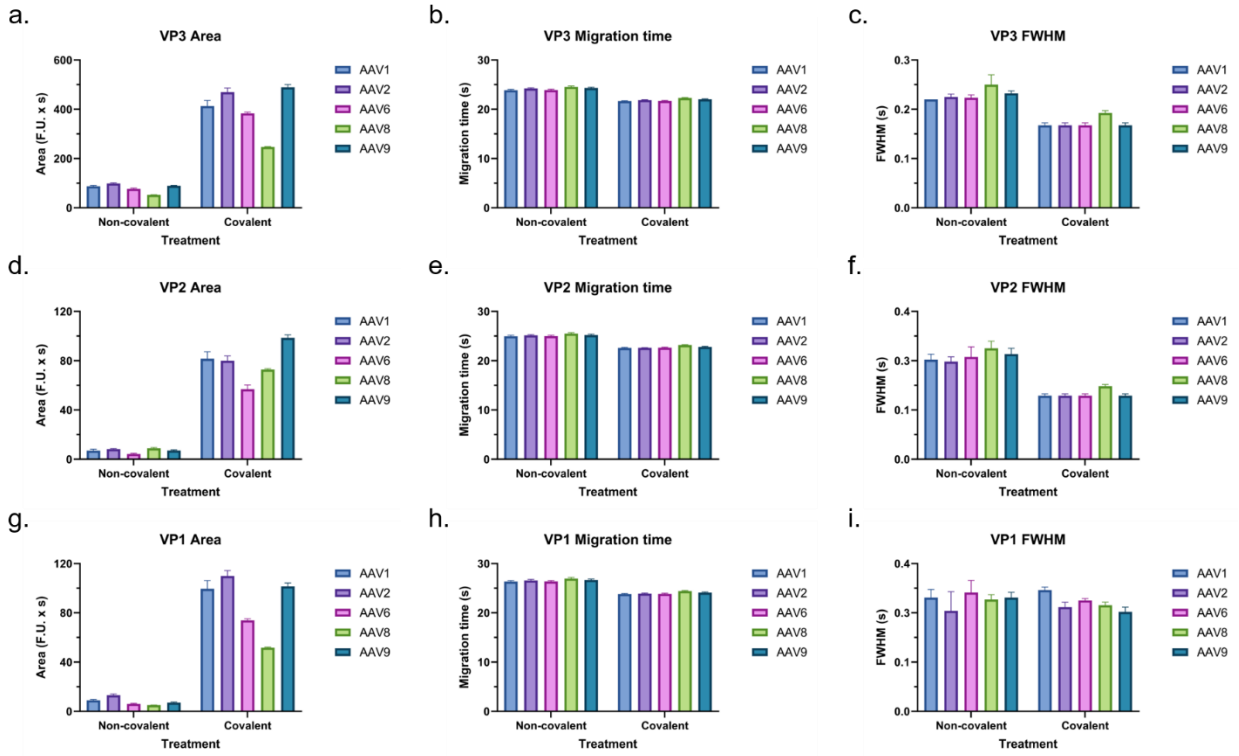

Figure S3.1: Comparison of VP properties across the two methods. (a) Area, (b) migration time, and (c) FWHM of the VP3 peak. (d) Area, (e) migration time, and (f) FWHM of the VP2 peak. (g) Area, (h) migration time, and (i) FWHM of the VP1 peak. Here, we highlight some of the key electrophoresis outcomes of the capsid VPs that we can assess with the proposed method and how they compare between the non-covalent and the covalent dyes. While the covalent peak areas were approximately 6 times larger than the non-covalent, the VP migration time and full-width half-maximum values were similar across the serotypes, suggesting a similar degree of protein denaturation and electrophoresis resolution for each sample.

Table S3.1: Sizing Accuracy of VPs using the Non-Covalent and Covalent Methods

Table S3.1: Sizing of the different capsid proteins and sizing accuracy (% CV) using the non-covalent and the covalent methods. The % CV values fall within the ≤ 20% cutoff value.

Serotype	Non-Covalent				Covalent			
	VP3 Size (kDa)	VP2 Size (kDa)	VP1 Size (kDa)	% CV	VP3 Size (kDa)	VP2 Size (kDa)	VP1 Size (kDa)	% CV
AAV1	72.9 ± 0.4	87.3 ± 0.5	107.3 ± 0.6	0.6	71.7 ± 0.3	85.1 ± 0.7	104.3 ± 0.6	0.6
AAV2	77.4 ± 1.2	89.8 ± 1.3	110.6 ± 1.6	1.5	74.4 ± 0.3	85.3 ± 0.5	105.9 ± 0.3	0.4
AAV6	73.9 ± 0.9	88.5 ± 1.1	108.4 ± 1.0	1.1	71.9 ± 0.2	85.4 ± 0.3	104.5 ± 0.5	0.4
AAV8	81.7 ± 1.0	94.9 ± 1.1	116.1 ± 2.0	1.4	79.8 ± 0.2	92.8 ± 0.2	113.1 ± 0.5	0.3
AAV9	78.8 ± 1.0	90.7 ± 0.9	111.9 ± 0.9	1.1	75.8 ± 0.2	87.3 ± 0.2	107.9 ± 0.9	0.3
	Average			**1.1**	**Average**			**0.4**

Table S3.2: Percentage Full Prediction Accuracy of AAV8 Reference Standard

Table S3.2: Percentage full prediction accuracy of AAV8 reference standard accuracy. The % CV values fall within the ≤ 20% cutoff value except for AAV1 57% full. The UV/Vis percentages of full were determined using UV/Vis (Stunner).

Serotype	UV/Vis % Full	Predicted % Full	Prediction Error (%)	% CV
AAV1	57	37.0 ± 14.3	20.0	38.6
	29	17.5 ± 3.3	11.5	19.1
AAV2	62	61.6 ± 4.7	0.4	7.6
	50	48.5 ± 3.7	1.5	7.7
	25	24.7 ± 4.5	0.3	18.1
AAV6	72	73.3 ± 12.6	1.3	17.2
	50	46.1 ± 4.8	3.9	10.3
	25	18.5 ± 3.2	6.5	17.5
AAV8	70	58.4 ± 6.5	11.6	11.2
	50	44.5 ± 3.7	5.5	8.2
	25	20.5 ± 2.6	4.5	12.8
Average			6.1	15.3

Figure S3.2: Comparison of UV/Vis (Stunner) and LabChip Percentage Full Predictions

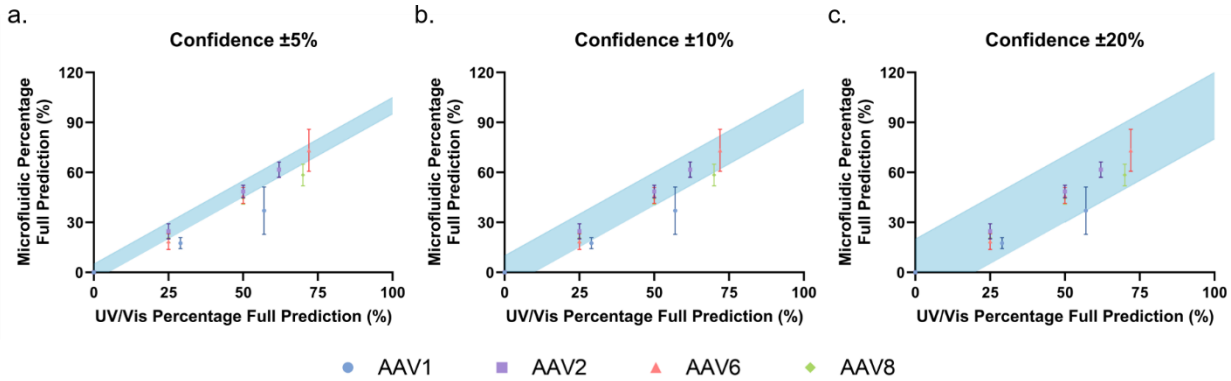

Figure S3.2: Comparison of the proposed method (LabChip) against UV/Vis (Stunner) at (a) ± 5%, (b) ± 10%, and (c) ± 20% confidence intervals with AAV1, 2, 6, and 8. To assess the performance of our microfluidic method against a more adopted method, we looked at how the microfluidic predictions (using AAV8 as the reference standard) compared to the UV/Vis (Stunner) estimations. For most serotypes, the microfluidic predictions were within ± 5-10% of the UV/Vis predictions, suggesting a close agreement between the methods. However, once again, AAV1 was the outlier, with one of the predictions being closer to ± 20%. While we believe this deviation can be attributed to a larger capsid ΔTm, we believe this highlights the potential and limitations of the technique, showing high accuracy for serotypes with lower ΔTm and a higher variation for those closer to the operating capsid ΔTm.

Chapter 4: Electrophoretic Microfluidic Characterization of mRNA and pDNA Loaded Lipid Nanoparticles

Chapter 4 has been submitted as an original research article.

Coll De Peña, A.,[1,+] Zimmer, D.,[2,+] Gutterman-Johns, E.,[3] Chen, N.M.,[3] Tripathi, A.,[1] Bailey-Hytholt, C.M.,[2] *Electrophoretic Microfluidic Characterization of mRNA and pDNA Loaded Lipid Nanoparticles*, **in review**.

[1]Center for Biomedical Engineering, School of Engineering, Brown University, Providence, RI, USA

[2]Department of Chemical Engineering, Worcester Polytechnic Institute, Worcester, MA, USA

[3]Department of Molecular Biology, Cell Biology, and Biochemistry, Division of Biology and Medicine, Brown University, Providence, RI, USA

[+]These authors contributed equally

4.1 Abstract

Lipid nanoparticles (LNPs) are clinically advanced non-viral gene delivery vehicles with a demonstrated ability to address viral, oncological, and genetic diseases. However, further development of lipid nanoparticle therapies requires new rapid analytical techniques to support their development and manufacturing. The method developed and described in this paper presents a new approach to rapidly and accurately analyze LNPs for optimized therapeutic loading by utilizing an electrophoresis microfluidic platform to analyze the composition of LNPs with different clinical lipid compositions (Onpattro®, Comirnaty®, and Spikevax®) and nucleic acid (pDNA and mRNA) formulations. This method enables the high throughput screening of LNPs using a 96- or 384-well plate with approximate times of 2-4 minutes per sample using a total volume of 11 μL. The lipid analysis requires concentrations approximately between 10^9 and 10^{10} particles/mL and has an average precision error of 10.4% and a prediction error of 19.1% when compared to using a NanoSight, while the nucleic acid analysis requires low concentrations of 1.17 ng/μL for pDNA and 0.17 ng/μL for mRNA and has an average precision error of 4.8% and a prediction error of 9.4% when compared to using a PicoGreen and RiboGreen assay. In addition, our method quantifies the relative concentration of nucleic acid per LNP. Utilizing this approach, we observed an average of 263 ± 62.2 mRNA per LNP and 126.3 ± 21.2 pDNA per LNP for the LNP formulations used in this study. We foresee the utility of this technique in the high throughput characterization of LNPs during manufacturing and formulation research and development.

4.2 Introduction

Lipid nanoparticle (LNP) delivery systems have accelerated to the forefront of therapeutic vehicles as a commercially viable, non-viral gene therapy that can be used to deliver small interfering RNA (siRNA), messenger RNA (mRNA), plasmid DNA (pDNA), and other nucleic acids (NA).[130–132] The range of NA types that can be loaded into an LNP increases their therapeutic versatility, a key consideration when determining the type of NA to be loaded into the LNP is the biological barriers that must be overcome before obtaining the desired therapeutic effect.[131,133–135] For instance, after cellular endocytosis of an LNP containing an RNA payload, endosomal escape and delivery into the cytosol results in translation.[136] Meanwhile, a DNA payload needs to be uptaken by the cell nucleus for transcription, increasing the diffusional barriers needed for NA uptake and subsequent protein expression.[133,136,137] In recent years, LNPs have gained worldwide attention as the mRNA carriers used in the Pfizer/BioNTech and Moderna COVID-19 vaccines.[134,138,139] Additionally, in 2018, Food and Drug Administration (FDA) approval was granted to the novel therapeutic Onpattro® (Patisiran), the first FDA-approved LNP carrying an siRNA payload for treating transthyretin-mediated amyloidosis.[140] The development and approval of Onpattro® demonstrated the safety and utility of LNPs as a carrier for gene therapy payloads, paving the way for research into various clinical treatment categories.[140]

LNP therapeutics efficiently load nucleic acids due to the incorporation of an ionizable lipid within the lipid composition.[139,141] The ionizable lipid typically has a pKa below 7,[142] resulting in protonation in low pH solutions, allowing it to complex with negatively charged nucleic acids. After LNP formation, a buffer exchange can be performed, resulting in a neutral particle and decreased cytotoxicity upon administration compared to cationic particles. The LNP shields the encapsulated nucleic acid from degradation following administration and helps cellular delivery

by promoting endocytosis.[143] Helper lipids, polyethelyne glycol (PEG) conjugated lipids, and cholesterol are incorporated in the lipid mixture for their structural and immunoregulatory properties.[141,144] While efforts are made to develop LNPs with low cytotoxicity, excess presence of lipids can still cause unwanted cytotoxic or immunogenic effects. Thus it is essential to minimize the lipid moieties required to deliver the nucleic acid payload.[145] Consequentially, there is a need for a new method to quickly and accurately optimize nucleic acid to lipid nanoparticle ratios. In addition to reducing cytotoxicity and immunogenicity, reducing the amount of lipid per dose is a more efficient use of resources in manufacturing, and alternative routes of administration, such as subcutaneous,[146] oral,[147] and inhalation,[148] could be further explored with more concentrated therapeutics. We have developed a method that addresses this need using electrophoresis microfluidics, allowing for rapid optimization of LNP formulations.

The current standard LNP characterization methods typically assess the physicochemical properties of the particles and the nucleic acid loading efficiency.[140] Size distribution of LNPs is measured using dynamic light scattering (DLS), and zeta potential distribution is measured using electrophoretic light scattering (ELS).[149–152] Additionally, multiangle light scattering (MALS)[149,153–156] and nanoparticle tracking analysis (NTA)[157,158] can be used for particle size and concentration measurements. To analyze the encapsulation efficiency (EE%) of nucleic acid-loaded LNPs, a PicoGreen™ and RiboGreen™ assay reagent is commonly used for DNA and RNA, respectively[159] (Table 4.1). To detect nucleic acid, the reagent binds to the nucleic acid, and the fluorescent signal per unit suspension volume containing LNPs is quantified. In addition to monitoring physicochemical and encapsulation properties, analytical techniques are being explored to further characterize lipid and nucleic acid content and quality in LNPs for clinical use.[153,160–162] To compare our technique with additional analytical methods being investigated, the

methods can be organized into categories including separation processes, online (sample flows from one analytical device to the next) and offline (sample is analyzed by one detector at a time and does not flow through) characterization, and destructive detectors (Table 4.1 and Table S4.1).

Chromatographic methods, such as size exclusion chromatography (SEC)[153] and reversed-phase (RP) liquid chromatography (LC), as well as field flow fractionation (FFF),[156,163] can be used for LNP separation prior to characterization.[160,164,165] After separation, non-destructive detectors can be attached in series; these include UV-Vis, refractive index (dRI), and Diode array detectors (DAD) to assess the concentration of both lipids and nucleic acids using methods such as RPLC. Destructive detectors can also be employed, including evaporative light scattering detectors (ELSD),[166] charged aerosol detectors (CAD),[160] and mass spectrometry (MS),[167] which have commonly been used in lipidomic analysis. Further, differential scanning calorimetry (DSC)[168,169] is often used to characterize the thermodynamic properties of lipids, and it can be used to assess the associated lipid phase transitions to characterize LNP-based carriers. A summary of the attributes of each of these techniques is provided in Table 4.1. In comparison, electrophoresis microfluidics provides several key attributes that are not present in any other system described (Table 4.1).

Electrophoresis microfluidics presents an attractive solution to overcome the time-limiting or personnel-intensive steps involved in the aforementioned techniques that is suitable for the analysis of the numerous samples produced during vaccine development.[170] It can achieve so through its precise control of fluids, high throughput capabilities, and rapid sample processing times.[6,51] The integration of electrokinetics into the system further increases its analytical potential by enabling the analysis of true plugs, or finite volumes, of sample through its fine control over molecular transport, increasing the resolution of the system by providing the user with control over

the width of the sample peaks.[171] Our microfluidic electrophoresis assay uses a 96- and 384-well plate-compatible platform to analyze LNPs, more specifically their lipid content and NA cargoes, including lipid concentration, and NA concentration and integrity by exploiting their electrophoretic and biochemical properties. To achieve the lipid and subsequent particle count characterization, a detergent, lithium dodecyl sulfate (LDS), that interacts with the LNPs and fluorescent dye of interest is mixed with the LNPs to enable its visualization, and the samples are then separated by their electrokinetic properties within the microfluidic system. In contrast, to analyze the nucleic acid content, a detergent, Brij® 58, is used to extract the nucleic acid load from the LNP, prior to its microfluidic electrokinetic fragment-based analysis. Electrophoresis microfluidics technique is the only method to provide high throughput analysis of lipids, nucleic acid, and NP count (Table 1). Additionally, it can be used to detect low NA concentrations (1.17 ng/μL for pDNA and 0.17 ng/μL for mRNA) without interference by lipids, strongly indicating this technique could transfer to exosomes (2.4-12 ng/mL in serum prior to ultracentrifugation or immunopercipitation[172]), liposomes, or other nonviral gene therapies. Such a streamlined method is needed as the high throughput and low sample volumes reduce the resources required for batch analysis. Knowledge of the concentration and particle counts can further optimize LNP formulations in new and innovative ways. The ability to detect the length of the nucleic acid on the LabChip platform screens for the integrity and purity of nucleic acids, as we have previously shown in the context of dsRNA detection.[173] In contrast, the lipid peak quantifies the amount of lipid present, but the results appear agnostic to the formulation. The convenience of this assay and its high throughput capabilities is ideal for many applications, including formulation development and manufacturing conditions where multiple batches need accurate and rapid assessment.

Table 4.1: Comparison of current techniques and our electrophoresis method used to characterize LNPs.

Type	Techniques	High throughput	Lipids	Nucleic acids	Size of LNP	NP counts	Charge of LNP	Bonding and thermodynamics
Non-destructive	UV-Vis	No	Yes*	Yes*	Yes†	No	No	No
	DAD	No	Yes*	Yes*	Yes†	No	No	No
	DLS	No	No	No	Yes	No	No	No
	ELS	No	No	No	No	No	Yes	No
	MALS	No	No	No	Yes	Yes	Yes	No
Destructive	MS	No	Yes	Yes	No	No	No	No
	CAD	No	Yes	No	No	No	No	No
	ELSD	No	Yes	No	No	No	No	No
Stand-alone	DSC	No	No	No	No	No	No	Yes
	EE% assay	Yes	No	Yes	No	No	No	No
	NTA	No	No	No	Yes	Yes	No	No
	Electrophoresis Microfluidics	Yes	Yes	Yes	No	Yes	No	No

Key: Used in combination with * RPLC, †SEC

4.3 Materials and Methods

4.3.1 Lipid Nanoparticle Development and Characterization

a. Lipid Nanoparticle Formulation Nucleic Acid and Lipid Solution Preparation

Figure 4.1a illustrates the overall process of developing and characterizing LNPs. All lipids used in this study were commercially purchased, including ionizable lipids DLin-MC3-DMA (MC3) (MedChemExpress, Monmouth Junction, NJ), ALC-0315, and SM-102 (Broadpharm, San Diego, CA), helper phospholipid distearoylphosphatidylcholine (DSPC), cholesterol, and 1,2-dimyristoyl-sn-glycero-3-phosphoethanolamine-N-[methoxy(polyethylene glycol)-2000] (ammonium salt) (PEG2000-PE) (Avanti Polar Lipids, Alabaster, AL). All stock lipids were suspended in 200-proof ethanol with stock concentrations of MC3, ALC-0315, SM-102, DSPC, cholesterol, and PEG2000-PE of 10 mM, 25 mM, 25 mM, 30 mM, 20 mM, and 1 mM, respectively. Green fluorescent protein (GFP) producing nucleic acids, gWiz-GFP pDNA, and DasherGFP

mRNA (Aldevron, Fargo, ND) were stored at -20°C. Dilutions of the nucleic acid in citrate buffer containing 5 mM sodium citrate, 5 mM citric acid, and 150 mM sodium chloride adjusted to pH 4.25 were made fresh upon LNP formulation. All other chemicals were purchased from MilliporeSigma unless noted otherwise.

b. LNP Formulation

Unless otherwise specified, the LNP formulation protocol described by Bailey-Hytholt et al.[159] was followed to synthesize the LNPs. Briefly, a lipid mixture was prepared using the Onpattro® (Patisiran), Comirnaty® (Pfizer/BioNTech COVID-19 vaccine), and Spikevax® (Moderna COVID-19 vaccine) lipid compositions. The Onpattro® composition consisted of MC3:DSPC:Cholesterol:PEG2000-PE at a molar ratio of 50:10:38.5:1.5, respectively. The same composition is used in the Spikevax® formulation except for using the SM-102 ionizable lipid instead of MC3. The Comirnaty® composition consisted of ALC-0315:DSPC:Cholesterol:PEG2000-PE at a molar ratio of 46.3:9.4:42.7:1.6, respectively. The LNP trade names are used throughout to describe the lipid composition used. The total lipid concentrations used in this study were 5 and 10 mM. The LNPs were formulated with target nitrogen to phosphate (N/P) ratios ranging from 3 to 75 to ensure the assay worked for a range of N/P ratios that are clinically relevant N/P ratios (3-8) to batches that may be outside the limits of detection of other assays (50-75). With this dynamic range of detection, we strongly believe this assay could have additional applications in other non-viral therapeutic areas. Nucleic acids were diluted in citrate buffer at a concentration resulting in the appropriate N/P ratio. A NanoAssemblr Ignite microfluidic platform (Precision Nanosystems, Vancouver, BC, Canada) was used to formulate the lipid nanoparticles using a 10 mL/min flow rate and a 3:1 aqueous to organic mixing ratio. Following formulation, the LNPs were immediately diluted with 1× phosphate-buffered

saline (PBS), and a buffer exchange was performed using 100 kDa Amicon® ultra centrifugal filters at 1000 × *g* (MilliporeSigma, Burlington, MA, USA). At least three exchanges with PBS were performed before collecting the LNPs and storing them at 4°C until use.

c. Physicochemical Characterization

After formulation, the LNPs were characterized for their hydrodynamic diameter, polydispersity, and zeta potential using a Zetasizer Nano (ZEN3600 software version 8.01.4906, Malvern Panalytical, Malvern, UK). To measure the LNP hydrodynamic diameter and polydiserpsity, the LNPs were diluted 40× in PBS and loaded into a semi-micro cuvette. To measure the LNP zeta potential, the LNPs were diluted 40× in ultrapure water and loaded into a folded capillary zeta cell.

To assess the particle concentration (particles/mL), the LNP samples were diluted 100-500× in PBS and analyzed in triplicate or quadruplicate using an NS300 NanoSight (Malvern Panalytical).

d. Loaded Nucleic Acid Encapsulation Efficiency

Encapsulation efficiency was measured using a Quant-iT™ PicoGreen™ and RiboGreen™ (Thermo Fisher Scientific, Waltham, MA, USA) assay for pDNA and mRNA, respectively. LNP samples were loaded on a fluorescence-capable 96-well plate. To half of the replicates, Triton X-100 was added, which breaks apart the LNPs, releasing the nucleic acid loaded. By quantifying both intact LNPs and ruptured LNPs, the amount of nucleic acid encapsulated and unencapsulated was calculated. A standard curve was made using the same nucleic acid loaded with and without Triton X-100 addition to correlate the fluorescence values to the concentration of loaded and unloaded NA. These results were compared to the NA assay results obtained from the LabChip GX II Touch (Revvity, Waltham, MA, USA) method, also highlighted in Figure 4.1a.

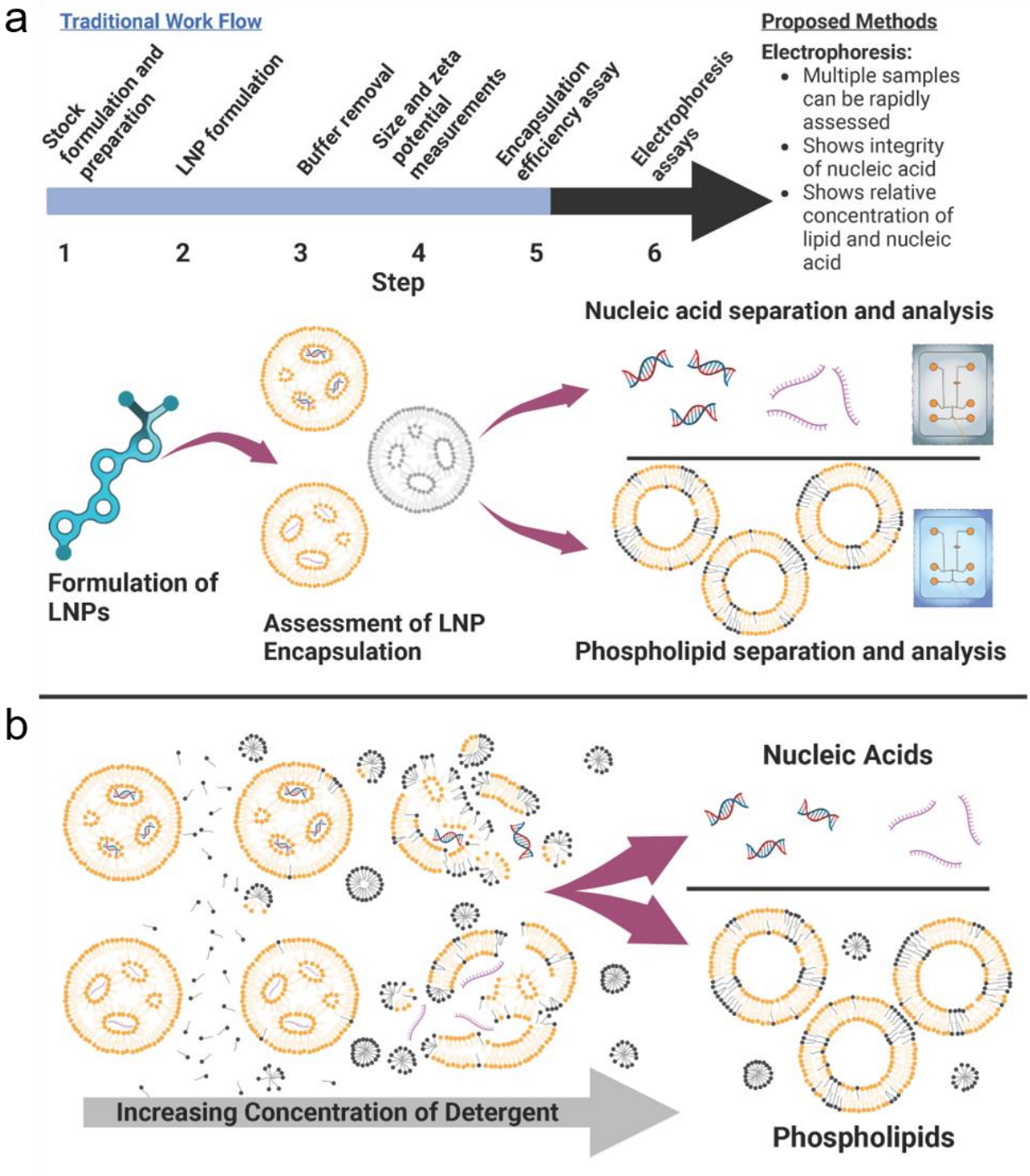

Figure 4.1: LNP formulation and breakdown for analytical development. (a) LNP formulation and assessment workflow. LNPs are first assembled using a microfluidic mixing platform, followed by physicochemical characterization and encapsulation efficiency measurements. The LNPs are then ruptured and assessed on a microfluidic chip, analyzing the relative concentration of nucleic acid to phospholipid. (b) Proposed mechanism of the breakdown of LNPs with increasing detergent concentration. LNP disassembly is important to reproducibly detect the nucleic acid and lipid content in the electrophoresis microfluidic technique. Figure created using BioRender.com.

4.3.2 Electrophoresis Microfluidic Analytical Method Development

a. LNP Preparation for Microfluidic Analysis

The analytical method developed and described here characterizes lipids and NAs independently and requires different sample preparations. LNP formulations were used unmodified, and a fluorophore was not included within the LNP composition to make this assay amenable for pharmaceutically relevant samples. Due to a fluorophore not being bound directly to LNPs, different detergents, including 0.5% (v/v) Triton X-100, 3% (w/v) lithium dodecyl sulfate (LDS), 3% (w/v) sodium dodecyl sulfate (SDS), and 6% (w/v) Brij® 58 were tested to treat the LNP samples prior to analysis to facilitate their on-chip dynamic fluorescence labeling (Figure S4.3). By heating the samples at 100°C for 5 min in the presence of these detergents, in this step, the LNPs break down and form a lipid complex during the sample preparation (Figure 4.1b), allowing the fluorophores (proprietary, provided by Revvity) to bind to the detergent dynamically on the chip.

The Triton X-100 and Brij® 58 treated samples did not produce a lipid peak, suggesting that an anionic detergent is needed for the dye to integrate into the lipid bilayer or that its migration is outside the window of the analysis potentially due to an opposite charge or a slower migration. In contrast, the LDS and SDS conditions produced a peak (Figure S4.3). In both cases using LDS and SDS, the control solution with PBS only and no LNPs present did produce a peak, but it was significantly smaller (7-fold smaller for LDS and 1.2-fold smaller for SDS) than when the LNPs were present. Given these results, the PBS control is subtracted as a blank, and the complexation can be observed with the SDS and LDS presence. Since the difference between the PBS peak and the LNP peak was more significant when using LDS, LDS was used for analysis as this more significant difference can result in a lower limit of detection as compared to the SDS treatment

(Figure S4.3). Therefore, based on these findings, 5 µL of the LNP sample was mixed with 7 µL of the selected pre-treatment sample buffer (Revvity) on a well plate, heated to 100°C for 5 min, mixed with 32 µL of nuclease-free water, and analyzed with a custom microfluidic chip and reagents (Revvity) for characterization of the lipids. Once the method was finalized, the individual lipids were analyzed and their electropherograms were compared to that of an LNP composed of the same lipids to characterize the potential labeling mechanism (Figure S4.4). We observed that DOPC/DSPC and PEG are important components for the ability to label the LNPs.

To extract the NA from the LNP for analysis, 6 µL samples were treated with 3 µL of 10× Revvity PicoRNA sample buffer, and 6 µL of 5% (w/v) LDS, 10% (w/v) aqueous Brij® 58 or 10% (w/v) Brij® 58 in formamide[170] (Figure S4.5). The LDS treatment failed to yield reproducible NA peaks. In contrast, both Brij® 58 solutions were successful at extracting the pDNA from the LNPs, but only the Brij® 58 formamide solution was able to fully extract the mRNA from the LNP, suggesting the formamide is needed to stabilize and protect the mRNA. Therefore, all NA experiments in this study were prepared by mixing 6 µL of the LNP sample with 20 µL of the Brij® formamide mixture and 3 µL of 10× Revvity PicoRNA sample buffer on a well plate. The samples were incubated at 70°C for 10 min, cooled for 5 min at 4°C, and analyzed using the PicoRNA chip and reagents (Revvity) and running the RNA Pico Sensitivity Assay on the platform. Only 11 µL of the sample are used for both assays at a concentration between 3×10^9 and 5×10^{10} particles/mL.

After this optimization, the final protocol involves two independent sample and chip preparations, one for the lipid and one for the NA analysis. The lipid analysis involves mixing 5 µL of the LNP sample with 7 µl of the pre-treatment sample buffer in a well plate, incubating it at 100ºC for 5 minutes, and adding 32 µL of nuclease-free water. In parallel, the custom LNP chip (Revvity) is

loaded with the provided custom LNP gel-dye (Revvity) that dynamically fluorescently labels the LNP sample for visualization as it electrophoretically migrates through the channels across the detection window. In contrast, the NA analysis involves mixing 6 µL of the LNP sample with 3 µL of 10× Revvity PicoRNA sample buffer and 6 µL of 10% (w/v) Brij® 58 in formamide in a well plate, incubating it at 70°C for 10 min and cooling it for 5 min at 4°C. In parallel, the PicoRNA chip (Revvity) is loaded with the PicoRNA gel-dye (Revvity) that dynamically fluorescently labels the NA for visualization as it electrophoretically migrates through the channels across the detection window. During both analyses, the samples are transferred from the well plate to the microfluidic chip using a vacuum pressure through the metal sipper. Once in the chip, the samples are transported and separated using electric fields.

b. Microfluidic Measurements and Analysis

The analytical platform LabChip GX II Touch (Revvity) was employed to control the pressure, electric fields, and optics in combination with the two different microfluidic chips used in this study. Before analysis, the microfluidic chips were loaded with a gel-dye mixture for the on-chip dynamic labeling of the samples and with a marker for the alignment of the electropherograms. LNP samples were treated to analyze either the lipids or the nucleic acid in a 96- or 384-well plate. The plates were transferred onto the LabChip platform and were individually loaded onto the chip via vacuum pressure applied by the system. Once in the chip, the samples were transported via pressure and electrokinetics across the chip and were electrophoretically separated via electrokinetics. The fluorescence intensity of the sample was recorded over time as the fluorescently labeled particles migrated over the optics probe. The resulting electropherograms were analyzed and interpreted using the LabChip GX Reviewer software version 5.4 (Revvity).

4.3.3 Statistical Analysis

Unless otherwise specified, all microfluidic experiments shown were conducted in triplicate technical repeat. The initial empty (non-NA loaded) and NA loaded LNP analysis shown comes from a single experimental repeat, which was in agreement with numerous iterations of the experiment not included in this study conducted at different concentrations. All other microfluidic experiments used a set of four dilutions of the original LNP samples, resulting in four experimental repeats per experiment, LNP type and batch. Results are reported as mean ± standard deviation. Statistical significance was determined using GraphPad Prism 9.4.1 (681) using an ordinary one-way ANOVA test followed by a Tukey post hoc test with a confidence interval of 95%; *$p<0.05$, **$p<0.01$, ***$p<0.001$, ****$p<0.0001$. The R^2 values of the linear relationships were determined using simple linear regression.

4.4 Results and Discussion

4.4.1 LNP Characterization using the Traditional Workflow

LNPs were formulated to test three different FDA-approved and pharmaceutically-relevant compositions as well as to compare two different nucleic acid types. Upon formulation using a NanoAssemblr Ignite, and following the workflow in Figure 4.1a, the individual LNP batches were analyzed for their hydrodynamic diameter, polydispersity, and encapsulation efficiency (Figure 4.2). At least two LNP batches were manufactured to test for batch-to-batch reproducibility or variability. The LNP hydrodynamic diameters quantified by DLS measurements were 79.8 ± 0.25 nm for the empty Onpattro® LNPs, 103.1 ± 2.23 nm for the pDNA Onpattro® LNPs, 92.5 ± 8.89 nm for the mRNA Onpattro® LNPs, 67.1 ± 1.68 nm for the mRNA Comirnaty® LNPs and 91.0 ± 1.89 nm for the mRNA Spikevax® LNPs (Figure 4.2a and Figure S4.1). All LNPs resulted a low polydispersity index (Figure 4.2b), where the empty Onpattro® LNPs had an average of 0.222 ±

0.0107, the pDNA Onpattro® LNPs 0.154 ± 0.0121, the mRNA Onpattro® LNPs 0.243 ± 0.0190, the mRNA Comirnaty® LNPs 0.130 ± 0.0327, and the mRNA Spikevax® LNPs 0.247 ± 0.087. The EE% for the pDNA Onpattro® LNPs was 99.6 ± 0.23%, 96.1 ± 1.07% for the mRNA Onpattro® LNPs, 89.3 ± 3.14% for the mRNA Comirnaty® LNPs, and 92.3 ± 1.46 for the mRNA Spikevax® LNPs (Figure 4.2c). Uniform size distributions were observed for all LNPs (Figure S4.1). The zeta potential varied depending on the LNP formulation, with an average of -2.3 ± 1.69 mV for the empty Onpattro® LNPs, -25.3 ± 2.45 mV for the pDNA Onpattro® LNPs, 16.2 ± 2.62 mV for the mRNA Onpattro® LNPs, -5.3 ± 6.48 mV for the mRNA Comirnaty® LNPs, and 10.4 ± 7.25 mV for the mRNA Spikevax® LNPs (Figure S4.1). These results are in line with previously reported physicochemical characterization of nucleic acid containing LNPs.[174–177]

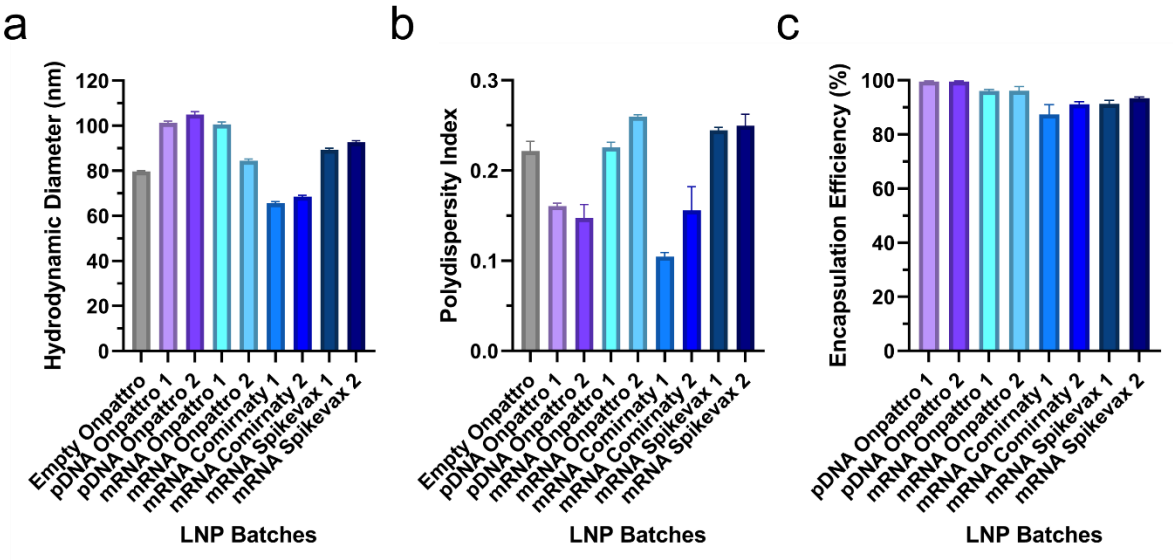

Figure 4.2: Physicochemical characterization of LNPs. (a) Average hydrodynamic diameter, (b) polydispersity index (PDI), and (c) Encapsulation efficiency (EE%).

4.4.2 Verification of LNP Disassembly using Detergent and Heat Treatment

Next, the concentration-dependent interaction between detergent concentration and LNPs was investigated. These verification studies were performed to ensure our method was causing an adequate and consistent breakdown of LNPs, their lipid structures, and nucleic acid release for

accurate analysis (Figure S4.2). To assess this, pDNA Onpattro® LNPs were heated to 99°C, followed by cooling for 20-30 minutes, with the addition of SDS to a final concentration of 8 mM. SDS and LDS critical micelle concentration (CMC) is 8.3 mM[178] and 8.5 mM,[179] respectively, and approaching this concentration is important to get an adequate breakdown of the LNPs. The results showed an adequate breakdown of the LNPs with two distinct size distributions forming, indicating micelles (from ~5 to ~15 nm) and larger aggregates (from 200 nm to 1000 nm) with no size overlap with the original LNP peak (~100 nm) (Figure S2C). This demonstrates sufficient detergent interaction with the phospholipids and is confirmed by the repeatable results observed throughout our experimental procedures.

4.4.3 LNP Characterization using the Traditional Workflow

While the initial lipid concentration used during LNP assembly are known and determined by the user, after LNP formation through microfluidic mixing, a buffer exchange is performed to remove the ethanol and change the pH to neutral. During this buffer exchange process, there can be a significant loss of material, including lipids, nucleic acid, and LNPs. Mousli et al. reported a 45-68% yield for the different ionizable/cationic lipids, pegylated lipids, helper lipids, and cholesterol components.[180] Currently, there is no quick and streamlined method to determine lipid loss in the LNP formulation process. Therefore, tracking lipid concentrations throughout the formulation process to use as a reference standard is not feasible. In contrast, quantifying the nucleic acid recovery can be calculated using the PicoGreen™ and RiboGreen™ reagents. For example, after LNP formulation in this study, an average mRNA yield of 85.2 +/- 9.0% was achieved as determined by the RiboGreen™ assay.

To accurately quantify the components of the LNPs, it is common practice in the field to rely on a reference standard. For our assay, lipid nanoparticle counts quantified using the NanoSight were

used as a reference for lipid characterization and the NA encapsulated concentration quantified using the PicoGreen™ and RiboGreen™ EE% assays. Particle concentration data obtained from the NanoSight is reported in Table S4.2. Another consideration is that LNPs can be loaded with varying NA strands per particle; therefore, it is crucial for an analytical method to quantify the amount of lipids or particles in a sample while being unaffected by the presence of NA. To assess the impact of NA presence on the lipid profiles, two sets of samples using the Onpattro® LNP formulation were analyzed at the equal particle concentrations of 3.83×10^{10} particles/mL to match the concentration of the empty LNPs based on the NanoSight analysis (Table S4.2). The samples included (1) an empty LNP and a pDNA loaded LNP and (2) an empty LNP and an mRNA loaded LNP. The empty and the NA loaded LNPs were normalized to the same concentration to enable a direct comparison between their lipid profiles. Before analysis in the LabChip platform, 5 µL of the LNP sample was mixed with 7 µL of a 5% (w/v) LDS pre-treatment sample buffer (Revvity) on a well plate, heated to 100°C for 5 min, and mixed with 32 µL of nuclease-free water.

As seen in Figure 3, the negative control (N.C.), which contained only PBS, produced a sharp peak of area A_{LDS} at ~20 s resulting from the interaction of the LDS with the fluorescent dye. However, both sets of LNP samples yielded significantly broader peak areas. The broad peaks are representative of LDS/Dye complexed with LNP constituents (MC3, DSPC, Cholesterol, and PEG2000-PE), which migrate with different velocity and diffusion coefficients. The area of the peaks is proportional to the number of LNPs, n_{LNP}, per unit volume, V, as described by:

$$A_{LNP} = A_{LDS} + \beta \frac{n_{LNP}}{V}, \tag{4.1}$$

where β is a constant for a fixed dye and LDS concentration. Since fluorescence and peak area fluorescence are relative, it is important to note that β will be constant throughout the analysis of

a well plate but not across different days and chip preparations since fluorescence signals are susceptible to subtle changes to the environment of the microfluidic channel. The data suggest that LNP constituents are conserved for both empty and full LNPs. Moreover, despite a significant area difference between the negative control and the LNP samples, there was no statistical difference between the area yielded by each LNP sample from the two sets. This suggests that the presence of nucleic acid does not interfere with the microfluidic lipid analysis and that the NA does not appear to be binding to or sequestering the lipid structure based on the electropherogram profile.

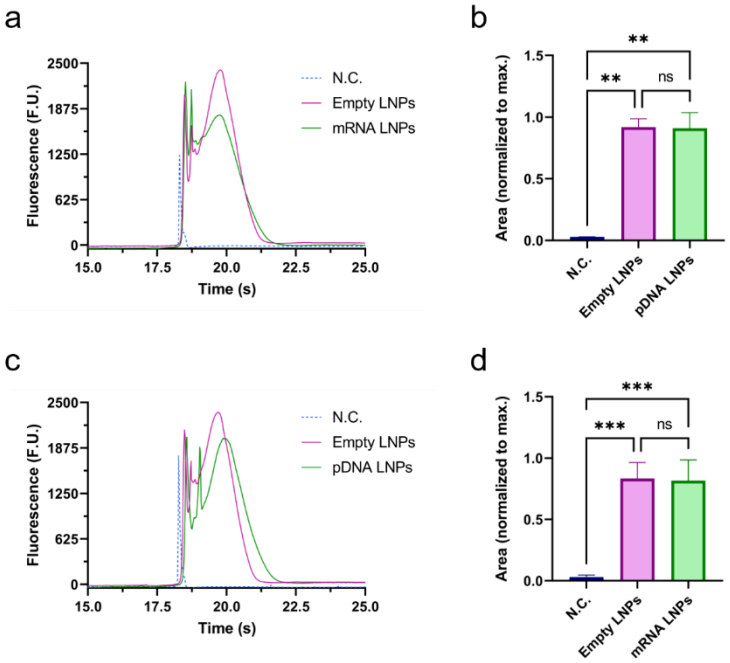

Figure 4.3: Microfluidic lipid analysis of LNPs. (a) Electropherograms and (b) peak area (normalized to the maximum peak area) yielded by the empty (Batch #1, Table S4.2) and diluted pDNA loaded (Batch #3, Table S4.2) Onpattro® LNP sample set. (c) Electropherogram and (d) peak area (normalized to the maximum peak area) yielded by the empty (Batch #1, Table S4.2) and mRNA loaded (Batch #2, Table S4.2) Onpattro® LNP sample set. A negative control consisting of PBS (blue), an empty LNP (pink) and a full pDNA (a and b) or mRNA (c and d) LNP sample (green) are displayed based on their electrophoretic analysis. All samples represented in this figure were diluted to a concentration of 3.83E10 particles/mL based on NanoSight results. Statistical analysis was performed using One-way ANOVA with a 95% confidence interval and Tukey post hoc analysis; *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

Next, dilutions of empty, pDNA, and mRNA loaded Onpattro® LNPs, as well as mRNA loaded Comirnaty® LNPs, and mRNA loaded Spikevax® LNPs were analyzed to assess the type of correlation between peak area and sample concentration. Figure 4 highlights the resulting lipid profiles at each dilution and the linearity in the response, with an average R^2 value of 0.93 ± 0.04. The strong correlation between LNP concentration and lipid peak area suggests the method can be used quantitatively to characterize samples of unknown concentrations. Regardless of the formulation, all samples produced peaks of equal migration times. In addition, the linearity in the response appeared to be independent of the LNP formulation, suggesting the platform is formulation agnostic. However, it is important to note that while the method does not rely on the formulation itself, based on experimentation, the linear range of quantitation is 3×10^9 and 5×10^{10} particles/mL. Below 3×10^9 particles/mL, it may be challenging to distinguish the sample from the blank. Above 5×10^{10} particles/mL, the response is no longer observed to be linear, likely due to the formation of different micellar structures induced by elevated lipid and detergent concentrations.[181]

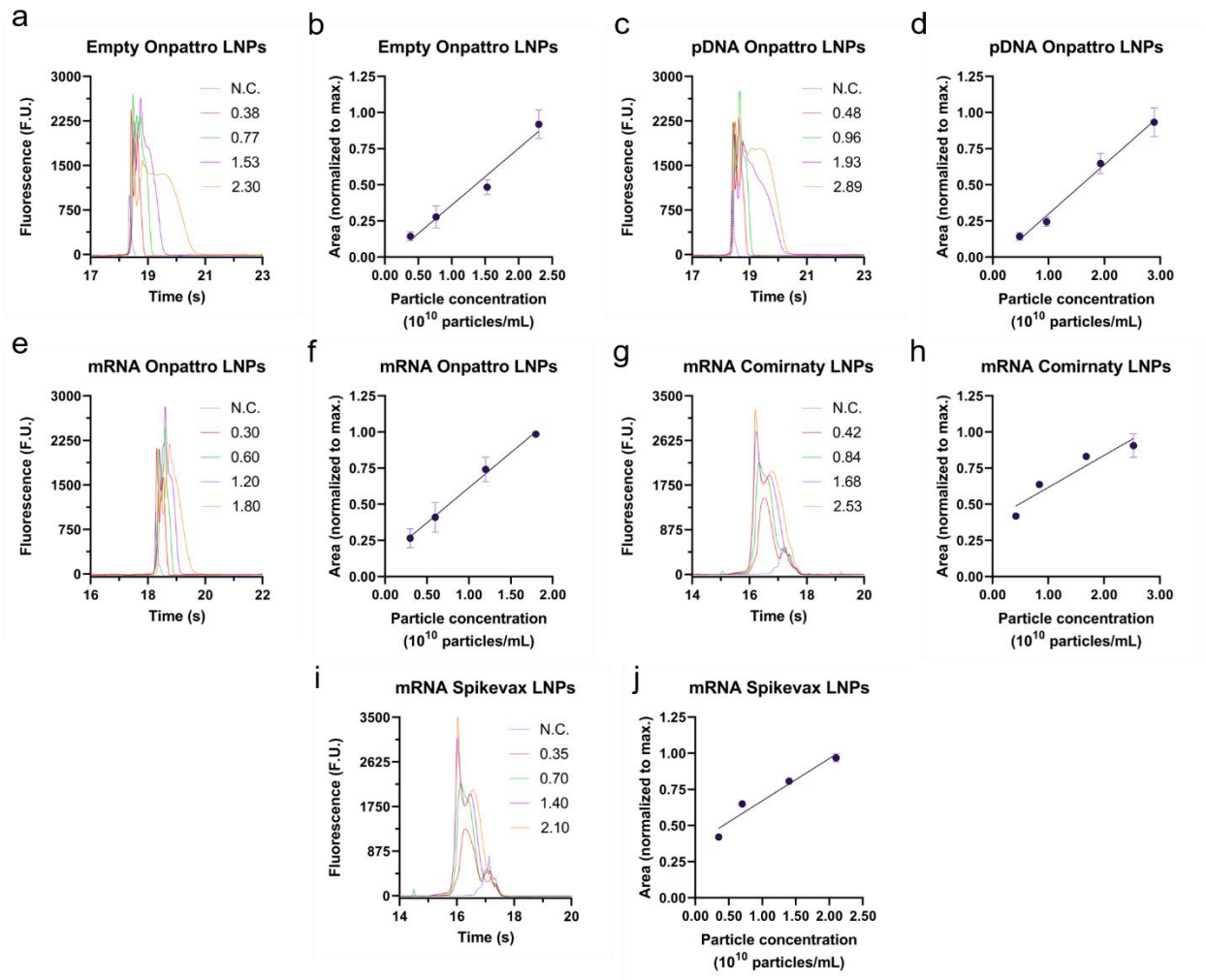

Figure 4.4: Assessment of linearity of LNP lipid response. (a) Electropherograms and (b) peak area (normalized to the maximum peak area) yielded by the empty Onpattro® LNP sample, with an R^2 of 0.94. (c) Electropherograms and (d) peak area (normalized to the maximum peak area) yielded by the pDNA loaded Onpattro® LNP sample, with an R^2 of 0.97. (e) Electropherograms and (f) peak area (normalized to the maximum peak area) yielded by the mRNA loaded Onpattro® LNP sample, with an R^2 of 0.96. (g) Electropherograms and (h) peak area (normalized to the maximum peak area) yielded by the mRNA loaded Comirnaty® LNP sample, with an R^2 of 0.87. (i) Electropherograms and (j) peak area (normalized to the maximum peak area) yielded by the mRNA loaded Spikevax® LNP sample, with an R^2 of 0.94. The values on the legend represent the particle concentration (10^{10} particles/mL). The average R^2 for all plots was 0.93 ± 0.04, suggesting a strong correlation between the particle lipid concentration and the peak area.

The linear correlation between peak area and concentration suggests our high throughput microfluidics method can be used as a measurement technique. More specifically, Figure 4.4 shows

that the number of LNPs per unit volume can be estimated, provided a standard of known concentration is used as a quality control indicator, using the following expression:

$$\frac{n_{LNP}^{sample}}{n_{LNP}^{standard}} = \frac{A_{LNP}^{sample} - A_{LDS}^{sample}}{A_{LNP}^{standard} - A_{LDS}^{standard}} \quad (4.2)$$

Our method can make the LNP manufacturing and quality control process more cost-efficient as the use of expensive, time-consuming techniques such as MALS and NTA (see Table 4.1) will not be required beyond the characterization of the standard if not provided.

4.4.4 Electrophoretic Method to Analyze Nucleic Acids

A key advantage of using non-viral delivery vehicles such as LNPs is that they are not as restricted in the type and length of NA that can be successfully loaded. As such, versatility is a crucial feature to consider when designing an analytical method for NA loaded LNP characterization. Considering the system requirements, the sample treatment was optimized to work with pDNA and mRNA fragments of 5,757 bp and 1,055 nt (gene length is 755 nt with an enzymatically added poly-A tail 200-400 nt long, assumed to be 300 nt for estimation purposes) in length, respectively. These two NAs were chosen to test a range of size and NA types in order to assess the capability of the same method being used for a range of payloads. As described earlier, after testing different detergents to extract the genetic material from the LNP, the most successful treatment was a heat incubation of the sample mixed with 10% (w/v) Brij® 58 in formamide.

To test the accuracy of the method, an initial experiment was conducted with two sets of Onpattro® samples; (1) an empty LNP containing no NA and a pDNA loaded LNP, and (2) an empty LNP containing no NA and a mRNA loaded LNP sample. Figure 4.5 shows that the empty LNPs did not produce a peak in either set of samples, while both the pDNA and mRNA loaded samples did. Moreover, given that the samples are electrophoretically separated based on size at a rate

dependent on their structure, we observed significantly different migration times for pDNA (t_{pDNA} ~34.5 s) and mRNA (t_{mRNA} ~46 s). Experiments were also conducted to compare the LNP-extracted NA integrity, and for both cases, the migration profiles were identical to the stock NA samples.

Since the pDNA used in this study has a much larger MW, 3,569.3 kDa, than the mRNA, 339.2 kDa (based on assumed 300 nt poly-A tail), pDNA has a significantly higher electrophoretic mobility than mRNA within the system. The pDNA sample has an estimated concentration of 64.92 ng/µL, whereas the mRNA sample has an estimated concentration of 4.83 ng/µL (Figure 4.5). Yet despite having a 13-fold concentration difference, the pDNA area is only 2-fold greater than the mRNA area, suggesting mRNA fluorescently labels more efficiently with this dye. In practical terms, the implication of these findings are that different NA types can be analyzed, the samples are analyzed fragment-based, enabling sample integrity assessment and undesired fragment detection, and NA-specific standards need to be used to estimate concentration based on fluorescent labeling efficiency.

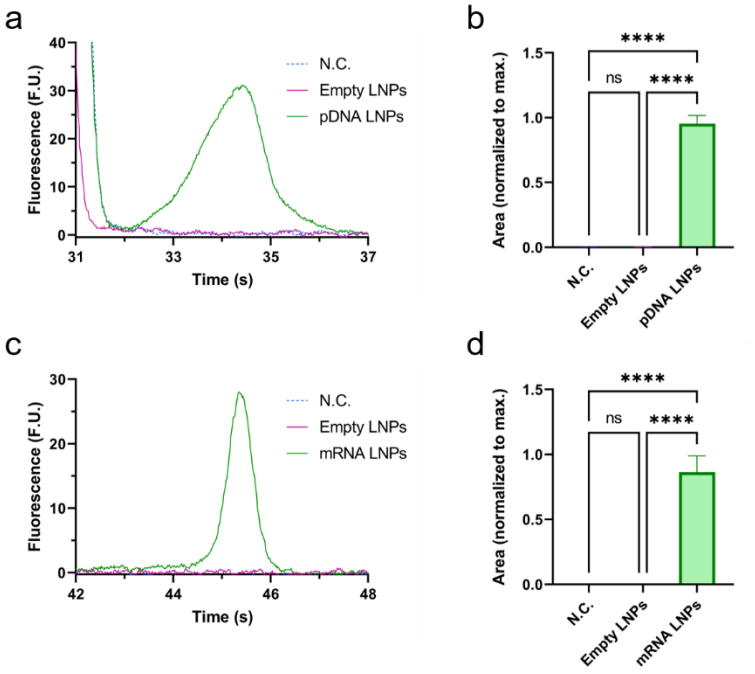

Figure 4.5: Microfluidic NA analysis of LNPs. (a) Electropherograms and (b) peak area (normalized to the maximum peak area) yielded by the pDNA loaded Onpattro® LNP sample set, where the first peak in (a) < 32 s is a lower marker used to align the electropherograms. (c) Electropherogram and (d) peak area (normalized to the maximum peak area) yielded by the mRNA loaded Onpattro® LNP sample set. A negative control consisting of PBS (blue), an empty LNP (pink) and a full pDNA LNP is displayed with an estimated concentration of 64.92 ng/µL (a and b) or mRNA LNP with an estimated concentration of 4.83 ng/µL (c and d) sample (green) based on their electrophoretic analysis. The NA concentrations were determined using PicoGreen™ and RiboGreen™. Statistical analysis was performed using One-way ANOVA with a 95% confidence interval and Tukey post hoc analysis; *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

Similar to the lipid dilution experiment, we then diluted pDNA loaded Onpattro® LNPs, mRNA loaded Onpattro® LNPs, mRNA loaded Comirnaty® LNPs, and mRNA loaded Spikevax® LNPs to assess the correlation between NA concentration and peak area (Figure 4.6). The area of the peaks, A_{NA}, are proportional to the concentration of NA, c_{NA}, as described by:

$$A_{NA} = \gamma c_{NA}, \tag{4.3}$$

where γ is a constant for a fixed dye concentration. The data suggests a strong correlation between concentration and peak area, with an average R^2 of 0.97 ± 0.02. The linearity in the response

suggests this method can be used quantitatively; this linearity was observed with a range 1:2 to 1:20 dilution for LNPs formulated between 5 and 10 mM. More specifically, Figure 4.6 shows that the NA concentration of an LNP provided a standard of known concentration is used as a quality control indicator, can be estimated using the following relation:

$$\frac{C_{NA}^{sample}}{C_{NA}^{standard}} = \frac{A_{NA}^{sample}}{A_{NA}^{standard}} \tag{4.4}$$

However, to use this analytical technique, the reference standard used to estimate the NA concentration needs to match the NA type of the LNP sample, given the difference in labeling efficiencies across NA type.

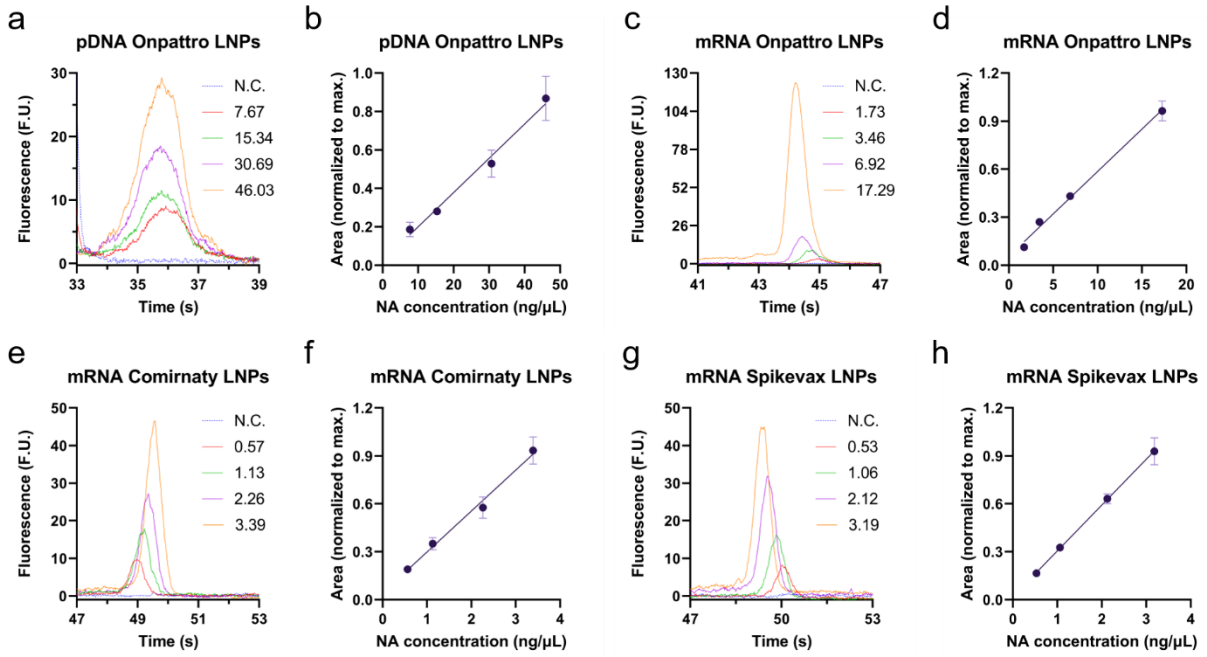

Figure 4.6: Assessment of linearity of NA response in LNPs. (a) Electropherograms and (b) peak area (normalized to the maximum peak area) yielded by the pDNA loaded Onpattro® LNP sample, with an R^2 of 0.94. (c) Electropherograms and (d) peak area (normalized to the maximum peak area) produced by the mRNA loaded Onpattro® LNP sample, with an R^2 of 0.99. (e) Electropherograms and (f) peak area (normalized to the maximum peak area) yielded by the mRNA loaded Comirnaty® LNP sample, with an R^2 of 0.96. (g) Electropherograms and (h) peak area (normalized to the maximum peak area) produced by the mRNA loaded Spikevax® LNP sample, with an R^2 of 0.98. The values on the legend represent the NA concentration (ng/μL) The average

R^2 for all plots was 0.97 ± 0.02, suggesting a strong correlation between the particle concentration or dilution and the peak area.

4.4.5 Validation of Analytical Method for Sample Concentration Predictions

In previous studies, we developed a microfluidic electrophoresis analytical method to characterize a viral gene delivery vehicle, adeno-associated virus (AAV).[109,182] Using a reference standard, we were able to predict the total protein and NA concentration and the percentage of full capsids present in an unknown AAV sample. The therapeutic relevance of this work is that it enables the monitoring of optimal LNP manufacturing and purification conditions to maximize the genetic load carried by the delivery vehicle. Given that the lipid and NA characterization methods showed a high correlation between concentration and peak area, our proposed analytical technique can be exploited to characterize samples of unknown concentrations (Equations 4.2 and 4.4). To do this, two independent batches were made for each formulation, and three dilutions of each batch were made using PBS to generate additional samples to test.

For the lipid analysis, Equation 4.2 was used to estimate the particle concentration. When the particle count was greater than 5×10^{10} particles/mL, which was determined to be the maximum loading capacity of the system, the stock was first diluted 2-fold in PBS before conducting the subsequent dilutions. The maximum loading capacity was determined when analyzing the mRNA loaded Onpattro® LNPs, which can be seen to exceed this threshold in Table 4.2. As a result, the concentrations above 5×10^{10} particles/mL were reported as outliers and excluded from the analysis. When excluding the errors associated with outliers, the % error associated with the lipid-based particle concentration prediction is 19.15% with a standard deviation of 9.32%. When separated by formulation, the pDNA loaded Onpattro® LNP formulation showed an average error of 24.53%, the mRNA loaded Onpattro® LNP formulation 23.76%, the mRNA loaded Comirnaty®

LNP formulation 15.20%, and the mRNA loaded Spikevax® LNP formulation 24.53%. Note that after determining the limit of the linear range of the lipid analysis with the mRNA Onpattro® LNPs, to overcome the higher LNP concentration of the pDNA Onpattro® LNPs, the sample and standard were first diluted 1:2 prior to analysis. While we did not use an orthogonal method to assess the accuracy of the NanoSight values, in our study, the precision error of the method 9.03 ± 3.86%, compared to 10.44 ± 8.73% for the LabChip method.

In contrast, the particle count range did not appear to impact the NA analysis, which used a different detergent in the sample buffer and different microfluidic chip, and therefore were not diluted. Using Equation 4.4, the NA concentrations of the samples were estimated using an non-encapsulated NA standard, which yielded an average error of 9.43% ± 7.26%. When divided into formulation, the pDNA loaded Onpattro® LNPs had an average error of 9.59%, the mRNA loaded Onpattro® LNPs 10.47%, the mRNA loaded Comirnaty® LNPs 6.98%, and the mRNA loaded Spikevax® LNPs 10.69%. Assuming a peak with an area of 1 F.U. × s can be observed, the NA limit of detection (L.O.D.) of the system was estimated from:

$$L.O.D. = \frac{C_{NA}^{sample}}{A_{NA}^{sample}}, \tag{4.5}$$

which suggests $L.O.D._{pDNA} = 1.17 \pm 0.13\, ng/\mu L$ and $L.O.D._{mRNA} = 0.17 \pm 0.02\, ng/\mu L$. However, the actual concentration ranges analyzed in this study were 7.67-64.92 ng/µL for encapsulated pDNA and 25 ng/µL for non-encapsulated pDNA, and 0.37-17.29 ng/µL for encapsulated mRNA and 25 ng/µL for non-encapsulated mRNA, all of which showed a linear response. A study by Raffaelle et al.[170] that also used the LabChip platform observed linearity from 2.5 to 15 ng/µL RNA, an LOD of 1 ng/µL for fragment-based detection although the RNA peak could be observed at lower concentrations and had a 14.1% coefficient of variation in peak area.

In contrast, it has been reported that RiboGreen and PicoGreen can have up to 30% difference based on components in solution[183] however no error tolerances are reported by ThermoFisher. Additionally, RiboGreen and PicoGreen assays are dependent on pipetting and the accuracy of the pipettes can cause compounding error, Eppendorf reports a tolerance for systematic error ranging from less than 1% up to 12% for its pipettes.[184] While we did not use an orthogonal method to assess the accuracy of the PicoGreen or RiboGreen values, in our study, the precision error of the method 4.45 ± 3.81%, compared to 4.84 ± 4.71% for the LabChip method (Table 4.2).

Table 4.2: Particle count predictions using an LNP standard for each LNP formulation and NA concentration predictions using a non-encapsulated NA standard for each NA type.

Sample	Measured Particle Concentration using NanoSight (particles/mL)	Particle Prediction using LabChip (particle/mL)	Prediction % Error	NA concentration using RiboGreen and PicoGreen assays (ng/μL)	NA Prediction using LabChip (ng/μL)	Prediction % Error
pDNA Onpattro® LNPs #1 (lipids analyzed at 1:2 dilution)	$9.64 \pm 0.33 \times 10^9$	$1.13 \pm 0.21 \times 10^{10}$	17.77	7.67 ± 0.90	10.30 ± 2.11	34.21
	$3.85 \pm 0.13 \times 10^{10}$	$5.11 \pm 0.55 \times 10^{10}$	32.71	30.69 ± 3.61	29.28 ± 3.88	4.59
	$5.78 \pm 0.20 \times 10^{10}$	$7.38 \pm 0.78 \times 10^{10}$	27.62	46.03 ± 5.41	48.08 ± 6.39	4.45
pDNA Onpattro® LNPs #2 (lipids analyzed at 1:2 dilution)	$8.59 \pm 0.69 \times 10^9$	$1.02 \pm 0.21 \times 10^{10}$	18.38	8.57 ± 0.08	8.79 ± 0.49	2.48
	$3.44 \pm 0.28 \times 10^{10}$	$4.16 \pm 0.27 \times 10^{10}$	21.09	34.30 ± 0.33	30.88 ± 1.59	9.97
	$5.16 \pm 0.41 \times 10^{10}$	$6.68 \pm 0.98 \times 10^{10}$	29.63	51.45 ± 0.50	50.49 ± 3.24	1.86
mRNA Onpattro® LNPs #1	$9.42 \pm 1.53 \times 10^9$	$1.19 \pm .017 \times 10^{10}$	26.11	1.73 ± 0.05	1.50 ± 0.01	13.06
	$3.77 \pm 0.61 \times 10^{10}$	$3.03 \pm 0.28 \times 10^{10}$	19.51	6.92 ± 0.21	6.29 ± 0.24	8.99
	$9.42 \pm 1.53 \times 10^{10}$	-	-	17.29 ± 0.53	18.08 ± 0.72	4.55
mRNA Onpattro® LNPs #2	$5.98 \pm 0.50 \times 10^9$	$3.90 \pm 0.42 \times 10^9$	34.84	0.48 ± 0.02	0.42 ± 0.01	13.74
	$2.39 \pm 0.20 \times 10^{10}$	$2.04 \pm 0.35 \times 10^{10}$	14.58	1.93 ± 0.07	1.77 ± 0.07	8.48
	$5.98 \pm 0.50 \times 10^{10}$	-	-	4.83 ± 0.18	5.50 ± 0.19	14.02
mRNA Comirnaty® LNPs #1	$4.21 \pm 0.27 \times 10^9$	$3.66 \pm 0.27 \times 10^9$	13.09	0.57 ± 0.05	0.65 ± 0.13	14.29
	$1.68 \pm 0.11 \times 10^{10}$	$1.64 \pm 0.06 \times 10^{10}$	2.60	2.26 ± 0.20	2.20 ± 0.16	2.64
	$2.53 \pm 0.16 \times 10^{10}$	$2.13 \pm 0.05 \times 10^{10}$	15.53	3.39 ± 0.30	3.23 ± 0.09	4.69
mRNA Comirnaty® LNPs #2	$4.06 \pm 0.33 \times 10^9$	$2.93 \pm 0.13 \times 10^9$	27.80	0.41 ± 0.01	0.44 ± 0.02	7.64
	$1.62 \pm 0.13 \times 10^{10}$	$1.51 \pm 0.02 \times 10^{10}$	6.83	1.63 ± 0.04	1.81 ± 0.03	11.04
	$2.44 \pm 0.20 \times 10^{10}$	$1.82 \pm 0.15 \times 10^{10}$	25.36	2.45 ± 0.06	2.49 ± 0.09	1.63
mRNA Spikevax® LNPs #1	$3.51 \pm 0.22 \times 10^9$	$2.58 \pm 0.13 \times 10^9$	26.39	0.53 ± 0.01	0.52 ± 0.03	2.28
	$1.40 \pm 0.09 \times 10^{10}$	$1.27 \pm 0.02 \times 10^{10}$	9.13	2.12 ± 0.03	1.95 ± 0.00	8.35
	$2.10 \pm 0.13 \times 10^{10}$	$1.65 \pm 0.01 \times 10^{10}$	21.51	3.19 ± 0.04	2.62 ± 0.13	17.64
mRNA Spikevax® LNPs #2	$2.28 \pm 0.17 \times 10^9$	$1.91 \pm 0.46 \times 10^9$	16.41	0.37 ± 0.01	0.39 ± 0.01	5.46
	$9.12 \pm 0.70 \times 10^9$	$1.04 \pm 0.01 \times 10^{10}$	14.22	1.48 ± 0.05	1.27 ± 0.02	14.14
	$1.37 \pm 0.10 \times 10^{10}$	$1.36 \pm 0.03 \times 10^{10}$	0.24	2.22 ± 0.08	1.86 ± 0.03	16.30
% Error Average	9.03	10.44	19.15	4.45	4.84	9.43

% Error Standard Deviation	3.86	8.73	9.32	3.81	4.71	7.26

Note: The standards and samples were different dilutions of the stock sample, and each sample set is an independent batch. The NanoSight and RiboGreen/PicoGreen concentrations were determined for one dilution of these samples and converted based on the dilution factor, the NanoSight and RiboGreen/PicoGreen values presented here are, therefore, not considering the errors associated with each method. The electrophoresis microfluidics predictions were made by extrapolating the concentration of the "unknown sample" based on its concentration relative to that of the standard. The % errors were calculated by dividing the absolute value of the difference over the actual concentration. The values excluded from the particle count predictions, "-", had a particle concentration that surpassed the linear range of the assay and were therefore excluded from the analysis (the linear range was determined after this data had been collected).

As an alternative to using a non-encapsulated NA standard for the concentration prediction, in the case where a non-encapsulated standard is not available and/or the NA concentration of the LNP standard is known, the use of an LNP NA standard was explored (Table S4.3). Like for the lipid analysis, for all formulations, the 0.2 dilution of the respective formulation batch was used as the standard, which was treated following the NA extraction protocol and analyzed on the same well plate as the LNP samples. Equation 4.4 was used once again to estimate the NA concentration of the samples, which yielded an average error of 10.22 ± 9.62%, regardless of NA load type. More specifically, the pDNA loaded Onpattro® LNPs had an average error of 12.50%, the mRNA loaded Onpattro® LNPs 9.73%, the mRNA loaded Comirnaty® LNPs 9.97%, and the mRNA loaded Spikevax® LNPs 8.69%. The similarity in the overall errors between the two methods using different standards further demonstrates the robustness of using this technique for NA containing LNP characterization, enabling the estimation of the NA load of an LNP with either an encapsulated or a non-encapsulated NA standard.

Using the concentrations of NA, C_{NA}, the molecular weight of the different NA, MW, and the particle counts, n_{LNP}, we obtain a ratio of the nucleic acid to LNP, $\frac{NA}{LNP}$, as highlighted in Figure 4.7 and as described by:

$$\frac{NA}{LNP} = \frac{\frac{C_{NA} \times N_A}{MW}}{n_{LNP}}, \quad (4.6)$$

where N_A is the Avogrado constant. The mRNA construct used has an estimated size of 339.1 kDa and 755 nt before enzymatic tailing (calculations assumed poly-A tail to be 300 nt) and a linear structure. In comparison, the pDNA has an approximate size of 3,569.3 kDa with 5,757 bp and a circular structure. The LNP samples used in this study ranged in size from 103.1 ± 2.23 nm for pDNA and 83.5 ± 12.99 nm for mRNA (Figure 2a). Using Equation 4.6, the pDNA loaded Onpattro® LNP formulation was estimated to contain an average of 134.3 ± 16.4 and 168.4 ± 13.6 nucleic acids per LNP with a LabChip estimate of 119.9 ± 29.5 and 132.8 ± 11.3, respectively for batches 1 and 2. The mRNA loaded Onpattro® LNP formulation had an average count of 326.0 ± 54.0 and 143.2 ± 13.1 nucleic acids per LNP with a LabChip estimate of 296.7 ± 101.7 and 171.6 ± 25.6 (these LabChip estimates only have two experimental repeats since the third was excluded for being above the maximum loading capacity of the chip), respectively for batches 1 and 2. Despite having a smaller LNP hydrodynamic diameter than the other two lipid compositions, the Comirnaty® formulation encapsulated an average nucleic acid molecules per LNP of 238.4 ± 25.8 and 177.6 ± 15.2 nucleic acids per LNP with a LabChip estimate of 273.6 ± 37.8 and 240.0 ± 26.2, respectively for batches 1 and 2. The Spikevax® mRNA formulation had 268.7 ± 17.0 and 287.9 ± 24.4 nucleic acid per LNP with a LabChip estimate of 303.3 ± 46.6 and 273.7 ± 78.5, respectively for batches 1 and 2. Among the different pDNA loaded LNP formulations, the average ratio was 151.4 ± 18.6 NA/LNP with a LabChip estimate of 126.3 ± 21.2, while among the different mRNA loaded LNP formulations, the average ratio was 241.0 ± 60.3 NA/LNP with a LabChip estimate of 263.0 ± 62.2. There are limited NA/LNP estimations reported in the literature and the ones available used different formulations and NA combinations than the one used in this study.

However, acknowledging those differences, we are including the values found in the literature for reference. The Leslie group[185] reported labeling ~100 siRNA/LNP using the formulation of KC2/Chol/DSPC/DSPE-PEG2000, with an average size of ~90 nm, N/P = 3, and using siRNA as their load. In contrast, Li et al.[186] reported ~2 mRNA/LNP with 40-80% empty LNPs depending on the assembly conditions using the formulation MC3/Chol/DSPC/DMG-PEG2000, with an average size of ~120 nm, using an mRNA 1,929 nt in length. They also show that at different N/P ratios and PEG concentrations, these values ranged from 5-15 mRNA/LNP when they decreased the mRNA length to 996 nt.[186] Interestingly, Henrickson et al.[187] reported that at N/P = 1 and N/P = 6, using a formulation of KC2/Chol/DSPC/PEG, their siRNA LNPs (> 50 nm in diameter) did not contain any empty LNPs. Using a similar method of analysis (NTA and RiboGreen), Staufer et al.[188] found that 54 ± 5.6 miRNA molecules per synthetic exosome, this is without the use of an ionizable lipid. Overall, these findings further support the dependence of NA loading with formulation and NA length, in addition to assembly parameters and pH, and the need for the described electrophoresis technique to characterize NA and LNP content in the manufacturing process.

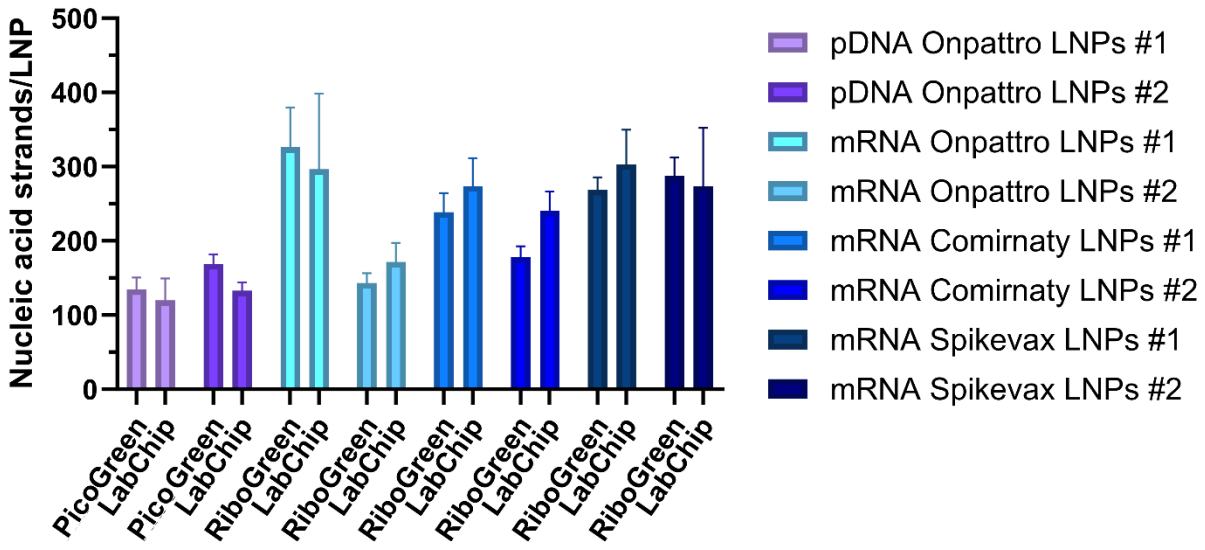

Figure 4.7: Estimation of nucleic acid strands per LNP. Comparison of the ratio is provided using the data obtained from the PicoGreen™ or RiboGreen™ with the NanoSight particle concentration, and the error of the method was determined using error propagation of the standard deviation yielded by a minimum of three replicates of each method, and with the NA and LNP concentration predictions using the LabChip electrophoresis microfluidics technique, where the error bars come from independent samples being analyzed. Nucleic acid strands per LNP estimations were obtained using Equation 4.6. Statistical analysis was performed using a Two-Way ANOVA with Šidák post hoc analysis and a 95% confidence interval. No statistical significance was observed between the PicoGreen/RiboGreen and the LabChip values.

4.5 Conclusions

Lipid nanoparticles have emerged as a gene delivery vehicle with great therapeutic potential, capable of addressing viral, oncological, and genetic diseases. However, given its recent widespread use, analytical tools to support its growing manufacturing needs have not emerged at the same rate, leaving a significant requirement for rapid, high-throughput characterization techniques. In this study, we developed a microfluidic electrophoresis method capable of analyzing the lipid and nucleic acid profiles of LNPs. Once the individual components are analyzed, an LNP reference standard can be used to predict the lipid or particle concentration within a 19.15% prediction error, and a non-encapsulated NA standard can be used to predict the NA concentration within a 9.43% prediction error. For both analyses, only 11 µL of the sample is required at a

concentration range of 3×10^9 and 5×10^{10} particles/mL for the lipid analysis, and above 1.17 ng/μL for pDNA and 0.17 ng/μL for mRNA nucleic acid analysis. Combining the run time of both assays, the method only takes 2-4 min per sample to run.

Moreover, a key advantage of using this microfluidic electrophoresis method to characterize the NA load of LNPs is the ability to do so in a fragment-based manner, as opposed to the fragment-property independent results yielded by traditional encapsulation efficiency assays. Therefore, NA concentration and integrity can be assessed simultaneously. We anticipate this platform can be further assessed for the analysis of lipid degradation (i.e., acidic, alkaline, and oxidative). Given the increasing interest in non-viral carriers, we also anticipate this platform will be able to analyze other types of lipid vesicles, such as extracellular vesicles and exosomes. In addition to the individual particle count and NA information yielded by this method, the information obtained from the two independent assays can also be combined to estimate the NA/LNP ratio. The ability to rapidly assess the NA/LNP ratio can be of great value for formulation optimization, as it allows users to screen many formulations in a short period of time. To allow for more streamlined optimization of LNPs, we are currently investigating a machine learning (ML) model that takes the formulation concentrations and formulation results as inputs to predict LNP count. Although the current ML predictions are promising, with specific models achieving an error of 17%, larger data sets with a more comprehensive range of formulations are still needed to optimize the model. In its current state, the high throughput microfluidic electrophoresis technique proposed in this study can be used to rapidly and accurately estimate particle and nucleic acid concentrations using a reference standard that has been previously characterized. In the future, through the implementation of ML, our goal is to bypass the need for off-chip standard characterization, further streamlining the use of this method for LNP formulation optimization and sample characterization.

4.6 Conflict of Interest

The authors declare no conflict of interest. A.T. is a paid scientific advisor/consultant and lecturer for Revvity.

4.7 Acknowledgements

This work was in part supported by Revvity's research grant to Brown University and by Worcester Polytechnic Institute. The authors thank Dr. Hong Susan Zhou at WPI for use of the Zetasizer Nano. The NanoSight NS300 data reported in this publication was supported by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant number P30GM114750 and the Kilguss Research Core at Women & Infants Hospital of Rhode Island. The authors would like to thank Dr. Sunil Shaw and Sarah Longley from the Kilguss Research Core for their training and support with the NanoSight.

4.8 Supporting Information

Table S4.1: Analytical Methods for LNPs

Analytical methods	Breaks apart the LNP?
FFF	No
SEC	No
RPC	Yes
EE% assay	Yes
DLS and ELS	No
Electrophoresis Microfluidics	Yes

Figure S4.1: LNP Size Distribution and Zeta Potential

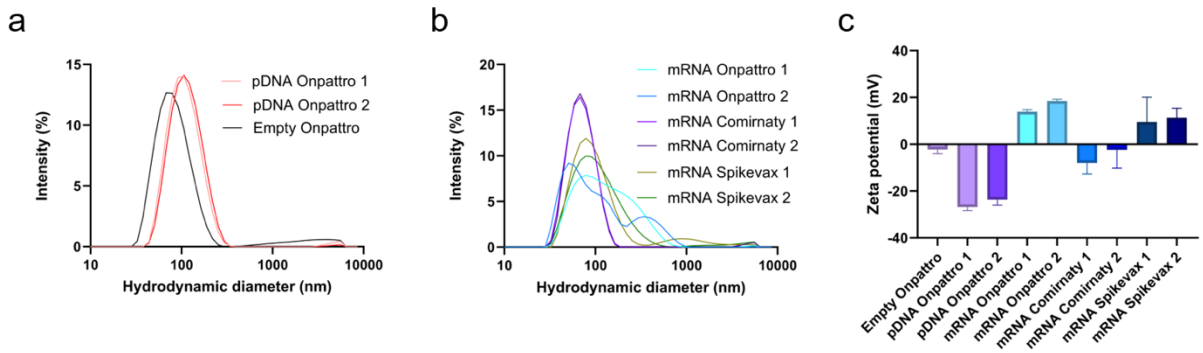

Figure S4.1: Hydrodynamic diameter size distributions of (a) empty (non-NA loaded) and pDNA Onpattro® LNPs, and (b) mRNA loaded LNPs. (c) Zeta potential of LNP batches. Number next to sample name indicate batch number.

Figure S4.2: Assessment of LNP Disassembly During Sample Preparation with SDS Detergent and Heat Treatments

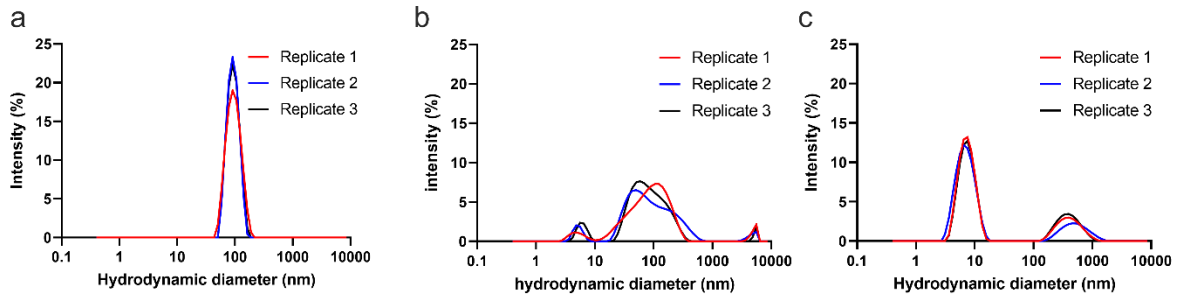

Figure S4.2: (a) Size distribution of LNPs without any SDS or heat treatment. (b) Size distribution of LNPs treated with 4 mM SDS and without heating after 25 minutes. (c) Size distribution of LNPs heated to 99°C for 30 minutes, followed by cooling at room temperature for 25 minutes and followed by the addition of 4 mM SDS.

Table S4.2: NanoSight LNP Particle Counts

NanoSight NS300 LNP particle counts of the samples used in the study. The empty Onpattro® LNPs, pDNA loaded Onpattro® LNPs, mRNA loaded Comirnaty® LNPs, and mRNA loaded Spikevax® LNPs were diluted 100× in PBS and analyzed in quadruplicate. The mRNA containing Onpattro® LNPs were diluted 500× in PBS and analyzed in triplicate. The LNP counts reported below were adjusted for the dilution factor.

Sample	Batch #	LNP count (10^{10} particles/mL)	Standard Deviation (10^{10} particles/mL)	% Error
Empty Onpattro® LNPs	1	3.83	0.40	10.5
pDNA Onpattro® LNPs	1	9.64	0.33	3.4
	2	8.59	0.69	8.0
	3	5.94	0.29	5.0
mRNA Onpattro® LNPs	1	9.42	1.53	16.3
	2	5.98	0.50	8.4
mRNA Comirnaty® LNPs	1	4.21	0.27	6.4
	2	4.06	0.33	8.2
mRNA Spikevax® LNPs	1	3.51	0.22	6.2
	2	2.28	0.17	7.6

Figure S4.3: Detergent Selection for Lipid Analysis

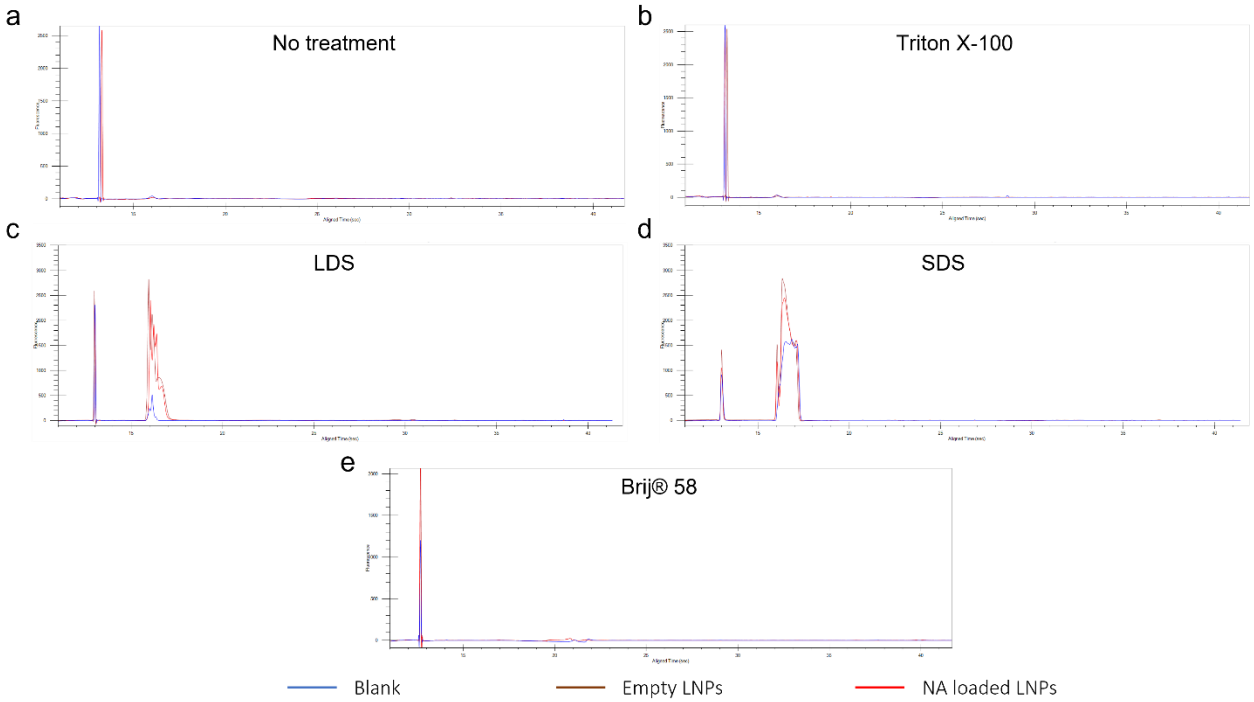

Figure S4.3: Electropherograms of the lipid analysis of the negative control containing only PBS (blue), pDNA loaded LNPs (red) and empty LNPs (brown) samples using (a) no detergents, (b) 0.5% (v/v) Triton X-100, (c) 3% (w/v) lithium dodecyl sulfate (LDS), (d) 3% (w/v) sodium dodecyl sulfate (SDS), and (e) 6% (w/v) Brij®58. All samples were analyzed by mixing 5 µL of sample (or PBS for blank) with 7 µL of PBS for no treatment and a detergent solution for the remaining samples. The samples were heated at 100 °C for 5 minutes and mixed with 32 µL of nuclease-free water prior to LabChip analysis. The first peak that is present in all conditions is a lower marker that was loaded directly onto the microfluidic chip for electropherogram alignment.

Figure S4.4: Individual Lipid and LNP Electropherograms using Microfluidic Lipid Analysis Method

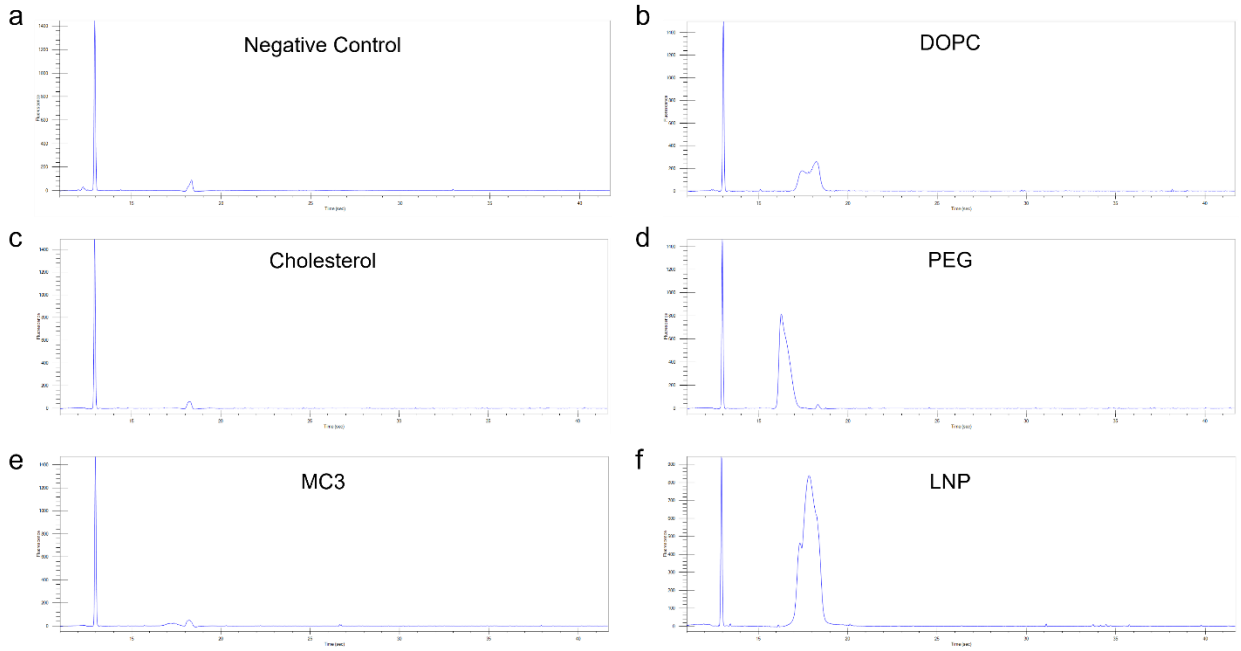

Figure S4.4: Electropherograms of the LDS-based lipid analysis of (a) the negative control containing only ethanol (which is the solution in which the individual lipids are diluted), (b) DOPC at 0.50 mM as an analogous substitute of DSPC, (c) cholesterol at 1.93 mM, (d) PEG at 0.08 mM, (e) MC3 at 2.4 mM, and (f) a DOPC-Cholesterol-PEG-MC3 LNP with approximately the same molar concentrations of the individual lipids. The individual lipid stocks are generally stored in PBS while the LNP samples are transferred to a PBS solution after formulation and dialysis. The electropherograms suggest that most of the signal in the LNP sample may be attributed to the DOPC and PEG present in the sample. However, because the ionizable lipid, in this case MC3, undergoes protonation during formulation, and the exact state of that individual lipid would be difficult to assess post formulation, the possibility that MC3 may also contribute to the signal cannot be ruled out.

Figure S4.5: Detergent Selection for Nucleic Acid Analysis

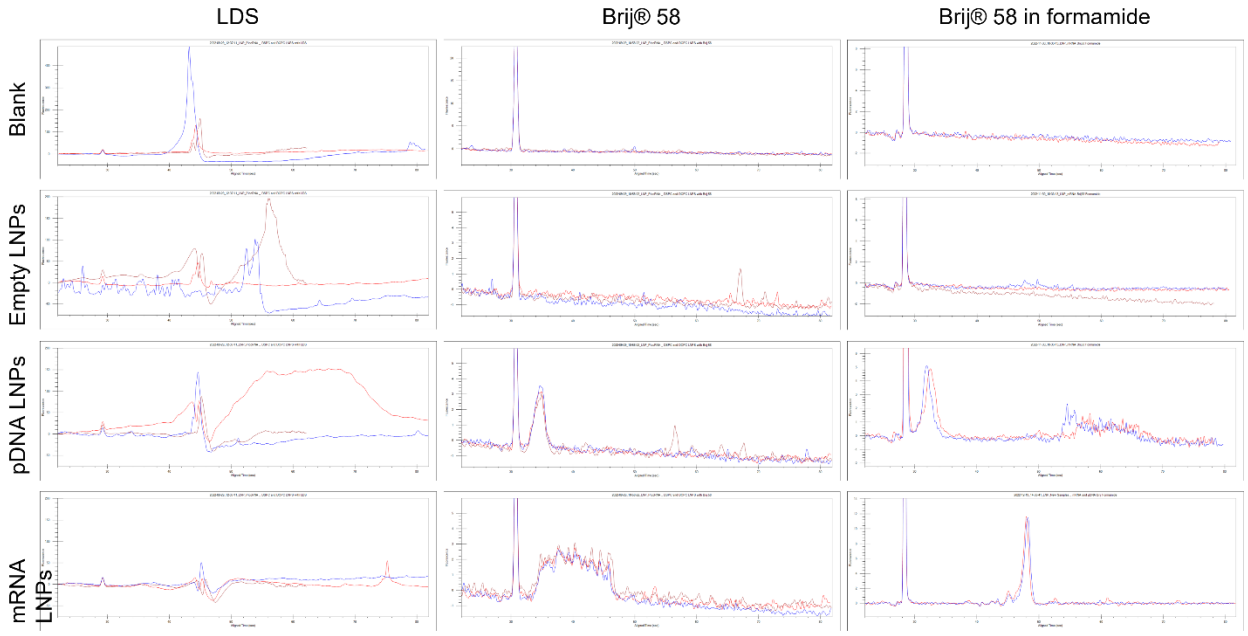

Figure S4.5: Electropherograms of the nucleic acid analysis for a negative control (PBS, row 1), empty Onpattro® LNPs (row 2), pDNA loaded Onpattro® LNPs (row 3) and mRNA loaded Onpattro® LNPs (row 4) treated with lithium dodecyl sulfate (LDS) (column 1), aqueous Brij®58 (column 2), and Brij®58 in formamide (column 3). The overlayed electropherograms (blue, red and brown) represent experimental repeats; 2-3 repeats were done for each condition. The sample treatments were conducted by mixing 6 µL of sample (or PBS for the blank), 3 µL of 10□ RNA sample buffer (Revvity, Waltham, MA) and 20 µL of the detergent being evaluated. Triton X-100 was also tested but clogged the chip and the run had to be aborted. A condition without detergent was also tested but it yielded no nucleic acid peaks and may have also affected the microfluidic chip. A lower marker was loaded directly onto the chip for electropherogram alignment, and it is the first peak present in all electropherograms. The only exception is the LDS treated empty LNP sample, as the baseline is heavily affected and reproducible peaks are not detectable.

Table S4.3: Comparison of Nucleic Acid Concentration Prediction using an LNP Standard and a Nucleic Acid Standard

An LNP standard (0.2 dilution of the stock) and a 25 ng/µL nucleic acid (pDNA for pDNA loaded LNPs and mRNA for mRNA loaded LNPs) standard were used to normalize the areas obtained from the electropherograms.

Sample	NA concentration using RiboGreen and PicoGreen assays (ng/µL)	NA Prediction using Electrophoresis Microfluidics and an encapsulated NA standard (ng/µL)	% Error	NA Prediction using Electrophoresis Microfluidics and a non-encapsulated NA standard (ng/µL)	% Error
pDNA Onpattro® LNPs #1	7.67	10.19	32.78	10.30	34.21
	15.34	*	*	15.51	1.08
	30.69	28.97	5.61	29.28	4.59
	46.03	47.57	3.34	48.08	4.45
pDNA Onpattro® LNPs #2	8.57	10.08	17.52	8.79	2.48
	17.15	*	*	14.96	12.80
	34.30	35.41	3.24	30.88	9.97
	51.45	57.90	12.54	50.49	1.86
mRNA Onpattro® LNPs #1	1.73	1.70	1.96	1.50	13.06
	3.46	*	*	3.07	11.32
	6.92	7.10	2.63	6.29	8.99
	17.29	20.38	17.90	18.08	4.55
mRNA Onpattro® LNPs #2	0.48	0.47	1.74	0.42	13.74
	0.97	*	*	0.85	12.21
	1.93	2.01	4.25	1.77	8.48
	4.83	6.27	29.88	5.50	14.02
mRNA Comirnaty® LNPs #1	0.57	0.72	27.89	0.65	14.29
	1.13	*	*	1.01	10.64
	2.26	2.46	8.95	2.20	2.64
	3.39	3.62	6.66	3.23	4.69
mRNA Comirnaty® LNPs #2	0.41	0.43	6.33	0.44	7.64
	0.81	*	*	0.82	1.24
	1.63	1.79	9.75	1.81	11.10
	2.45	2.46	0.26	2.49	1.50
mRNA Spikevax® LNPs #1	0.53	0.52	1.14	0.52	2.28
	1.06	*	*	1.05	1.15
	2.12	1.97	7.28	1.95	8.35
	3.19	2.65	16.68	2.62	17.64
mRNA Spikevax® LNPs #2	0.37	0.44	20.10	0.39	5.46
	0.74	*	*	0.65	12.19
	1.48	1.45	2.22	1.27	14.14
	2.22	2.11	4.69	1.86	16.30
Mean Prediction % Error			10.22		9.03
Error standard deviation			9.62		6.82

**Note: Since the 0.2 dilution was used in the cases where the LNP standard was used, there are no prediction values for those samples. Overall, the predictions using an LNP standard had a mean error of 10.26 ± 9.60%, while those using the NA standard yielded an average error of 9.06 ± 6.81%. While the NA standard predictions yielded slightly lower errors and standard deviations, the comparable results highlight the versatility and robustness of the method, enabling the estimation of the NA load of an LNP with either an LNP or NA standard.*

Chapter 5: A microfluidic electrophoretic dual dynamic staining method for the identification and relative quantitation of dsRNA contaminants in mRNA vaccines

---

Chapter 5 has been published as an original research article.

Coll De Peña, A.,[1] Li, N.,[1] Vaduva, M.,[1] Bwanali, L.,[2] Tripathi, A.,[1] *A microfluidic electrophoretic dual dynamic staining method for the identification and relative quantitation of dsRNA contaminants in mRNA vaccines*, **Analyst**, 148, 3758, 2023.

[1]Center for Biomedical Engineering, School of Engineering, Brown University, Providence, RI, USA

[2]Applied Genomics, Revvity, Hopkinton, MA, USA

5.1 Abstract

mRNA vaccines (*i.e.*, COVID-19 vaccine) offer various advantages over traditional vaccines in preventing and reducing disease and shortening the time between pathogen discovery and vaccine creation. Production of mRNA vaccines results in several nucleic acid and enzymatic by-products, most of which can be detected and removed; however, double-stranded RNA (dsRNA) contaminants pose a particular challenge. Current purification and detection platforms for dsRNA vary in effectiveness, with problems in scalability for mass mRNA vaccine production. Effectively detecting dsRNA is crucial in ensuring the safety and efficacy of the vaccines, as these strands can cause autoimmune reactions with length-symptom dependency and enhance mRNA degradation. We present a new microfluidics method to rapidly identify and quantify dsRNA fragments in mRNA samples. Our innovation exploits the differences in the dynamic staining behavior between mRNA and dsRNA molecules to detect dsRNA contaminants in a high throughput approach. The limit of detection of the system for dsRNA was estimated to be between 17.7–76.6 pg μL^{-1} with a maximum loading capacity of mRNA of 12.99 ng μL^{-1}. Based on these estimated values, our method allows for the detection of dsRNA contaminants present in percentages as low as 0.14–0.59% compared to the total mRNA concentration. Here, we discuss the molecular mechanism of the dynamic staining behavior of dsRNA and mRNA for two different stains. We believe our method will accelerate the mRNA vaccine development from initial development to quality control workflows.

5.2 Introduction

Vaccines have been a crucial public health measure in preventing and reducing diseases, morbidity, and mortality by millions each year.[189] Vaccine importance has been highlighted by the COVID-19 pandemic and the widespread use of newly developed mRNA vaccines. Despite early bottlenecks, much research has gone into nucleic acid-based vaccines due to their ability to provide precise targeting of the immune response and offer advantages in safety, efficiency, and specificity when compared to other vaccine platforms.[3,190] mRNA vaccines, specifically, offer benefits including extranuclear activity, which confers a minimal risk for subsequent random genome integration and insertional mutagenesis, more controlled expression of coded antigen, no inclusion of foreign genes, and the ability to be produced in a cell-free environment by in vitro transcription.[190,191] If this technology is harnessed for widespread and large-scale use, mRNA vaccine application could extend to cancer therapies, therapeutic protein replacement therapies, treatment of genetic diseases, and a wide variety of infectious diseases.[190,191] Recent advances in lipid nanoparticles (LNP) and mRNA technology over the years have increased the stability and efficacy of mRNA–LNP vaccines.[2,134,191]

However, contaminants such as leftover enzymes, free nucleotides, residual DNA, truncated RNA, and dsRNA are also possible during the synthesis of the mRNA.[3] These exogenous contaminants, specifically dsRNA, identified as a major contaminant, are potent pathogen-associated molecular patterns (PAMPs). The creation of dsRNA in the in vitro transcription (IVT) production of mRNA is believed to stem from the activity of T7 RNA polymerase.[192] When recognized, PAMPs can lead to the inhibition of translation and the enhancement of mRNA degradation. dsRNA contaminants can also activate pro-inflammatory cytokines associated with potential autoimmune reactions, including effects on the central nervous system.[3,193–195] Additionally, there appears to be a positive

correlation between dsRNA fragment length and the associated adverse effects,[196] suggesting that careful monitoring of both the presence and length of dsRNA fragments is required to ensure the safety of the vaccine.

Despite current purification platforms, there remains a lack of well-established manufacturing platforms for mRNA, thus numerous mRNA synthesis and purification techniques are generally combined.[191] The pharmaceutical industry often utilizes different forms of chromatography coupled with tangential flow filtration for purification due to its versatility, cost-effectiveness, selectivity, and, importantly, its ability to be upscaled for mass production of mRNA vaccines.[191] However, different forms of this technique come with their limitations. For example, ion pair reverse phase chromatography is an excellent purification method that removes dsRNA while maintaining a high yield, but it is very costly to scale and uses toxic reagents such as acetonitrile.[191,197] Conversely, anion exchange chromatography can be used to purify mRNA at a large scale cost-effectively, but it often requires denaturing conditions, tight control of temperature, and the use of potential chaotropic agents.[191] Lastly, high-performance liquid chromatography has been shown to remove dsRNA and other contaminants, resulting in a 10- to 1000-fold increase in protein production levels upon vaccination.[198] Still, there are issues with the purification scale-up process and mRNA stability.[197]

However, despite the presence of various purification platforms, some having evidence of dsRNA removal of over 90%,[199] the variability in quality and effectiveness of current methods call for quality assessment and quality control systems for detecting residual and harmful dsRNA fragments in mRNA vaccines. Currently preferred quality control detection assays for characterizing RNA transcripts include UV spectroscopy, fluorescence-based assays, immunoassays, and chromatography. However, these are severely limited by resolution, precise

handling, hazardous reagents, intense labor, need for antibodies, long run time, or inability to be scaled,[48,197,200–202] leaving a critical need for fast, high throughput, low-cost quality control systems.

The versatility of microfluidics allows for the fast development of novel analytical methodologies to monitor the effectiveness of the purification methods and ensure the safety of the mRNA vaccines. Furthermore, as a result of the miniaturization of traditional biomolecular analysis tools, the decrease in overall scales offers precise control of fluids, high-throughput capabilities, and rapid sample processing, making it capable of outperforming traditional technologies, often leading to lower-cost alternatives.[6,51] To obtain such fine control over the dynamics within the system, microfluidics is usually coupled with a driving force. Here, we use electrokinetics as the driving force as it exploits unique behaviors that would not be easily achievable at a larger scale. Here, we present a high-throughput microfluidic electrophoresis methodology that exploits the differences in fluorescent staining kinetics between dsRNA and mRNA to detect and characterize dsRNA contaminants in mRNA vaccines.

5.3 Materials and Methods

5.3.1 Sample and Sample Preparation

The ssRNA (catalog #N0364S) and dsRNA ladders (catalog #N0363S) used in this study were purchased from New England Biolabs (New England Biolabs, Ipswich, MA). The mRNA and dsRNA samples were custom ordered from Genewiz (Genewiz Genomics Headquarters, South Plainfield, NJ) and have 4001 nt and 4001 bp lengths, respectively. The sequence of the custom samples can be found in the SI (Supporting Information, Sequence).

5.3.2 Microfluidic System and Chip Reagents

A custom glass nucleic acid (NA) microfluidics chip (Revvity, Waltham, MA) with a metal sipper was interfaced with the high throughput LabChip GX Touch II platform (Revvity) for robotics motion, electrical, and flow controls. Two different far-red fluorescent stains (SYTO 61 as a DNA stain and RiboRed as an RNA stain; ThermoFisher Scientific Inc., Waltham, MA) were obtained from Revvity and utilized at varying concentrations (Fig. S5.1–S5.3). These stains were selected because dsRNA will have the uracil (or chemically modified uracil) bases that are characteristic of RNA but a structure more similar to DNA, both of which could influence the labeling behavior of dsRNA. Therefore, both stains were evaluated to investigate which nucleic acid type dsRNA simulated more closely in terms of fluorescent staining. Different concentrations of poly(N,N-dimethylacrylamide) (Revvity) were used, as specified throughout the study, to load the microfluidic channels after mixing with the fluorescent stain of interest. A custom lower maker (Revvity) was loaded onto one of the chip wells and injected using a combination of pressure and electrokinetics for electropherogram alignment.

5.3.3 Methods

Depending on the requirements of the experiments, the NA samples were diluted in nuclease-free water to achieve the desired concentration and maintain the total salt concentration <150 mM and were transferred onto a 96- or 384-well plate. The NA microfluidic chip was loaded with the gel-dye mixture of interest and the lower marker. The well plate and the chip were then transferred onto the LabChip platform for analysis. Once the optimal assay conditions were established, the sample plate was analyzed in a chip loaded with a SYTO 61 gel matrix, and upon completion, the same sample plate was analyzed using the same chip loaded with a RiboRed gel matrix, or vice versa; stain analysis order does not affect the results (Fig. 5.1). The statistical analyses included in

this study were conducted on GraphPad Prism 9.4.1 with a confidence interval of 95% where *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001.

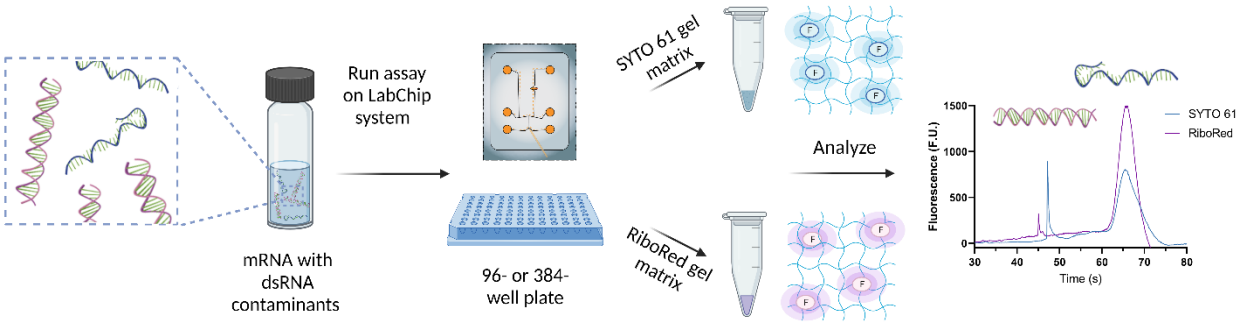

Figure 5.1: Sample preparation and analysis workflow for the detection of dsRNA contaminants in mRNA samples and vaccines. Figure assembled/created using BioRender.com.

5.4 Results and Discussion

5.4.1 Understanding the Differences between dsRNA and ssRNA

Fundamental experiments in this study were conducted with a dsRNA ladder to simulate dsRNA contaminants and an ssRNA ladder to simulate mRNA. The findings from these experiments were then used as a reference or starting point for the analysis of the custom-ordered long (4001 bp or nt) dsRNA and mRNA samples. Long dsRNA and mRNA molecules are not readily available, particularly highly purified, and limited information regarding the electrokinetic response of those samples is available in the literature.

The first step was to characterize the mobilities and fluorescent staining response of ssRNA and dsRNA with respect to heating (Fig. S5.1) and fluorescent stain type (SYTO 61 and RiboRed; Fig. S5.2). Generally, ssRNA molecules are heated before analysis to prevent the presence of multimeric forms and tertiary structures from interfering with their mobility.[203] However, upon analyzing both ssRNA and dsRNA, it was observed that while heat resulted in better-defined peaks for the ssRNA sample, it also denatured smaller strands of dsRNA, causing the fabrication of

dsRNA peaks. Acknowledging that numerous techniques exist for the analysis of mRNA molecules and that the focus of this study is to characterize dsRNA contaminants, heating was excluded to optimize the conditions for dsRNA identification. Next, Fig. S5.2 shows the fluorescent response of a single dsRNA and ssRNA fragment at various SYTO 61 and RiboRed concentrations. Based on the dsRNA signal output, the optimal concentrations were 2.92% for SYTO 61 and 15.00% for RiboRed. Lastly, Fig. S5.3 shows a snapshot of migration times as a function of gel concentration. The optimal gel concentration, 3%, was picked based on fragment migration time and peak width to ensure assay resolution. When the gel was lowered to 3%, the new stain optimal concentrations were determined to be 1.17% and 15% for SYTO 61 and RiboRed, respectively, to achieve better repeatability dynamic staining under the new gel concentration.

Then, the dsRNA and ssRNA ladders were analyzed using the microfluidic electrophoresis chip platform, and the following electropherograms were yielded (Fig. 5.2A). No significant difference in mobility was observed between the two RNA types (Fig. 5.2B and Table S5.1) within this size range (50–500 bp or nt) at our gel concentration of 3% poly(*N*,*N*-dimethyl acrylamide). However, during preliminary experiments with a 6% gel, ssRNA had slower mobility than dsRNA on average (Fig. S5.1). To better understand these findings, we investigated further the main sample factors that affect their mobility (μ) within our microchannel loaded with a semidilute polymer solution. For simplification purposes, this relation has been represented in terms of net drag (D) and net charge (q) as: $\mu \sim \frac{net\ charge}{net\ drag}$.

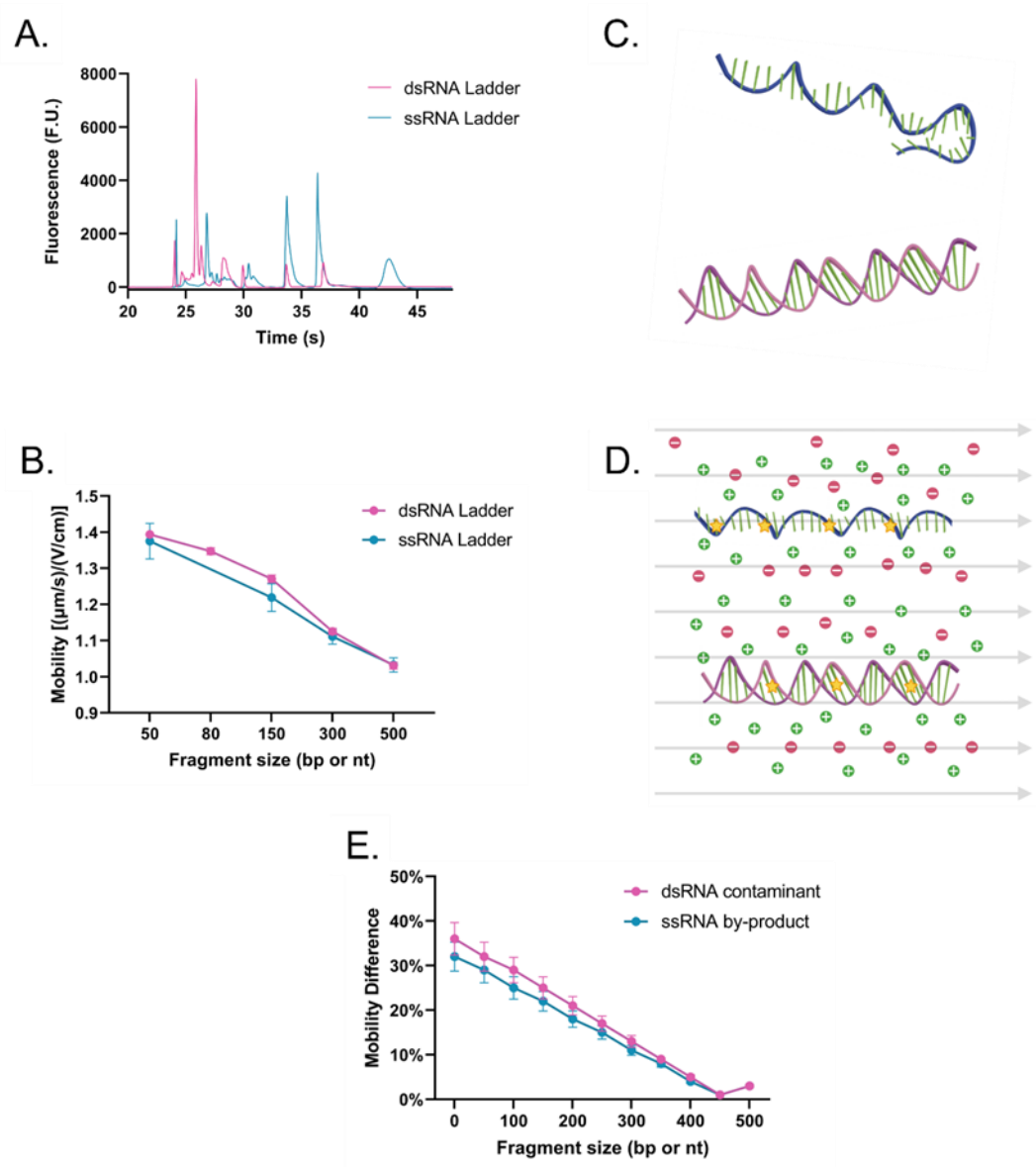

Figure 5.2: (A) Electropherogram (the leftmost peak of each ladder represents the system lower marker; some additional minor peaks surrounding the lower MW species were determined to be low MW impurities in the ladder, but since only the main peaks are being utilized, it is not expected to affect the analysis) and (B) the average mobility of dsRNA (pink) and ssRNA (teal) ladders using SYTO 61. (C) Schematic representation of an ssRNA or mRNA (teal) and a dsRNA (pink) molecule in the absence of external factors, and (D) in the presence of an electric field (grey arrows), which causes them to stretch and reorient, ions (green plus signs and red minus signs) and fluorophores (yellow stars) to highlight the complex interactions experiences by the analytes. (E) Differentiation potential of the method between dsRNA (pink) and ssRNA (teal) contaminants from a 500 nt long vaccine based on the experimentally obtained mobility at different fragment sizes normalized to the 500 nt ssRNA fragment mobility.

ssRNA is generally more prone to forming tertiary structures than dsRNA (Fig. 5.2C); however, during electrophoresis, the electric field can stretch and reorient nucleic acid molecules in the direction of the field, especially in the case of shorter chains (Fig. 5.2D).[204] As a result, it is likely that the mobility of ssRNA will be less affected by its tertiary structure in the case of shorter chains, while longer ones may see an impact. In contrast, due to the presence of a second strand in dsRNA, under conditions where ssRNA is straightened, dsRNA may have a bulkier conformation and experience more drag than ssRNA. Persistence length (l_p), which represents the stiffness of a molecule, also plays a key role in sample fragment mobility. It has been reported that dsRNA l_p ranges from 62–72 nm,[205,206] while that of ssRNA ranges from 0.5–1.27 nm.[207,208] Generally, a lower persistence length will enable the molecule to migrate more easily through the matrix, giving the advantage to ssRNA.[209] However, when the focus is shifted to the contribution of the charge on the mobility instead since the charge density of dsRNA is greater than that of ssRNA, dsRNA is expected to migrate faster through the system.[210]

Given that at lower gel concentrations within the 50–500 base range, there is no observable difference between the two molecule types, but that at higher gel concentrations there is, the data suggests that at low gel concentrations, shorter ssRNA fragments can be straightened and orient to themselves to the electric field and through the contribution of its persistent length can migrate at the same velocity as dsRNA, despite the latter being favored by having a charge that is approximately twice in magnitude. However, the data also suggests that at higher gel concentrations, despite its higher flexibility, through a combination of its tertiary structure and having less charge than a dsRNA molecule of the same length, ssRNA migrates slower, expanding the current knowledge of how these molecules migrate in comparison to each other in microfluidic electrokinetic systems.[53,211]

However, despite the ability to induce a mobility difference via gel concentration, it was determined that mobility alone could not be used to identify dsRNA contaminants. During the synthesis of IVT-mRNA, both truncated ssRNA fragments (<mRNA length) and, over time, degraded mRNA fragments (<mRNA length) can be present in the sample, which, despite a mobility difference can overlap with the dsRNA contaminants (≤mRNA length; Fig. 5.2E). This led us to explore alternative methods of differentiation and identification.

The main difference that stood out upon closer examination of the data was that despite both ladders being expected to have similar total RNA concentrations, most fragments in the ssRNA ladder fluoresced brighter than those in the dsRNA ladder, like in earlier findings (Fig. S5.2). Due to the overlaps in structure between dsRNA and dsDNA and in biochemistry between dsRNA and ssRNA, we explored both SYTO 61 (DNA stain) and RiboRed (RNA stain) for the fluorescent visualization of our molecules. Experiments were conducted using the previously determined concentrations where a dsRNA and an ssRNA ladder were analyzed separately after exposure to each stain. Again, we noticed that the overall fluorescence of the ssRNA peaks was higher than those of the dsRNA peaks, but more interesting than that, we noticed that the way the ladders responded to the fluorescent stains was opposite to each other. While dsRNA ladder fragments generally produced lower degrees of fluorescence than ssRNA ladder fragments when labeled with both stains, we noticed the following trend: when dsRNA was exposed to SYTO 61, the dynamic staining was more effective than that offered by RiboRed while the ssRNA stained better with RiboRed (Fig. 5.3 and Tables S5.2, S5.3).

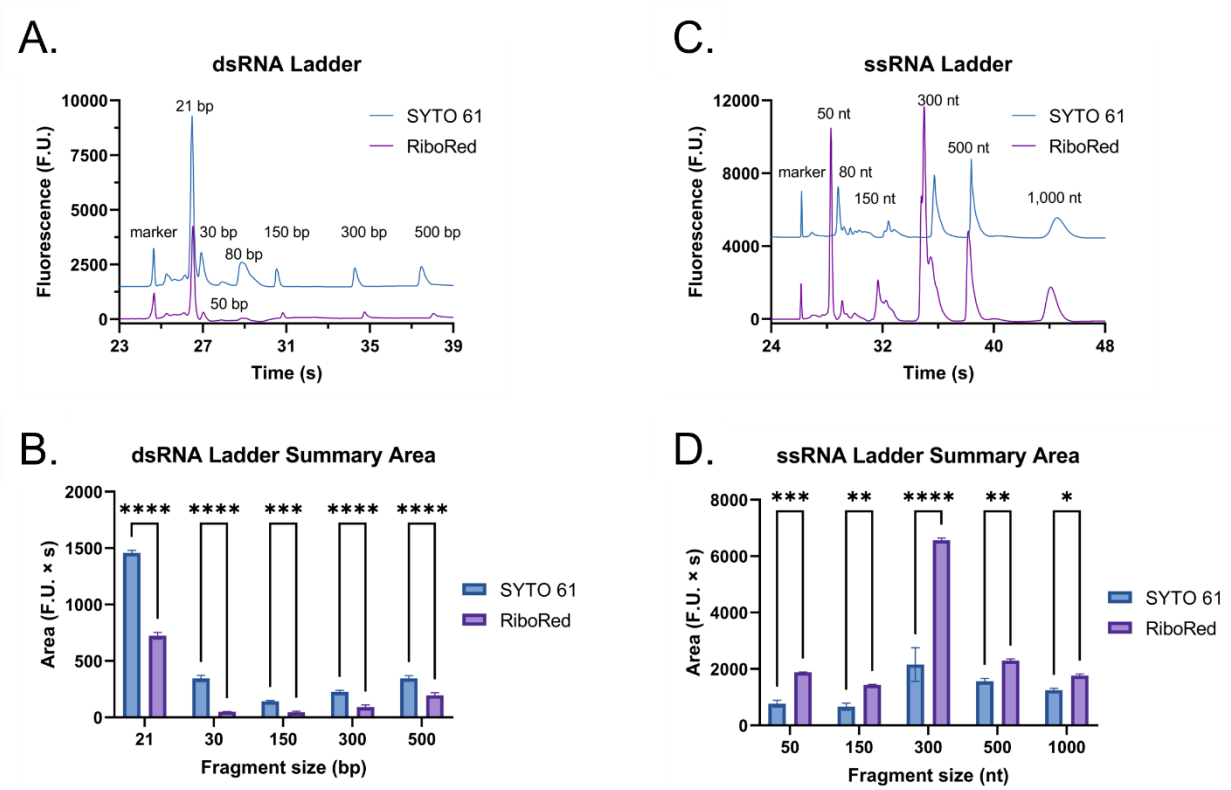

Figure 5.3: Fluorescent staining differences between a dsRNA and an ssRNA ladder using SYTO 61 (blue) and RiboRed (purple). (A) Electropherograms offset by 1500 fluorescent units and (B) the average area produced by the dsRNA ladder with each stain. (C) Electropherograms offset by 4500 fluorescent units and (D) the average area produced by the ssRNA ladder with each stain. In the bar graphs, "*" indicates the significance level in the difference (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$), while "ns" indicates no statistical difference.

In fact, when comparing the areas yielded by each stain, on average, the dsRNA ladder yielded an area 56% greater with SYTO61 than with RiboRed, while the ssRNA ladder yielded an area 118% greater with the RiboRed dye. Interestingly, while the same trend was observed in the peak area and peak height for each fragment size, 10/10 peak areas showed a statistical difference, whereas only 7/10 peak heights (Fig. S5.4) did. This suggests that peak area may be a better indicator when assessing the effect of different fluorescent stains. This is the first evidence we obtained that suggested differential dynamic staining kinetics across the two types of ribonucleic acid molecules could be used to identify dsRNA in mRNA vaccines.

5.4.2 Assessing Translation of Findings to Long dsRNA and mRNA Fragments

To assess whether the findings made using the shorter dsRNA and ssRNA fragments were also applicable to long fragments (4001 bp or nt), we analyzed a long dsRNA and a long mRNA molecule with the same sequence (SI, sequence) using our system. Interestingly, unlike with shorter fragments, Fig. 5.4 suggests that there is a significant difference in mobility between long dsRNA and mRNA molecules. This is supported by our hypothesis that shorter ssRNA fragments are more easily straightened than longer fragments. As a result, longer ssRNA or mRNA fragments will have a slower mobility since they experience a higher conformation-induced drag with no charge density increase relative to a dsRNA of equivalent length.

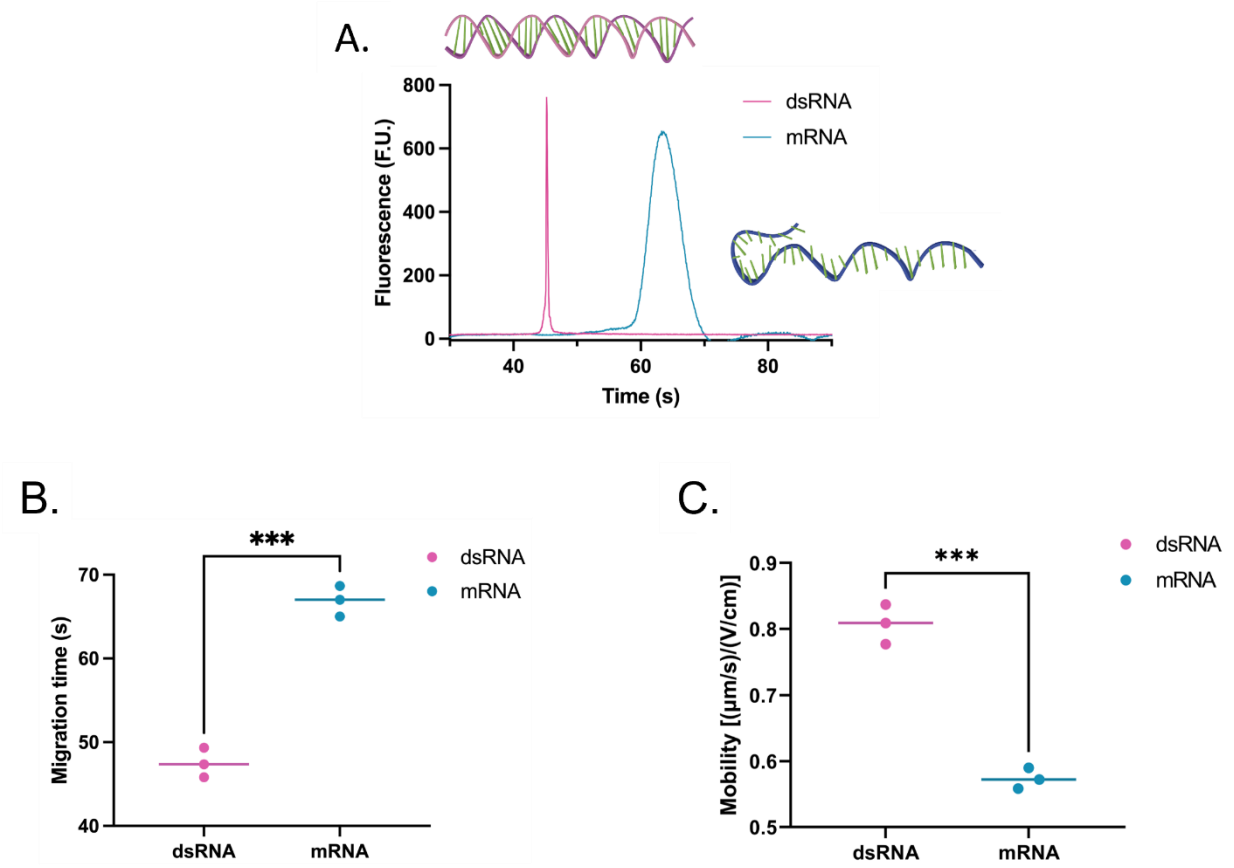

Figure 5.4: (A) Electropherogram, (B) migration, and (C) mobility of a 4001 bp dsRNA (pink) and 4001 nt mRNA (teal) fragment using SYTO 61. The migration times include data from three independent runs, each with two to three repeats. In the comparisons, "*" indicates the significance level in the difference (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$).

Moreover, while the dsRNA sample had a concentration of 2.60 ng μL^{-1} and the mRNA 5.90 ng μL^{-1}, instead of showing an increase in area equivalent to the change in concentration (2.3 times) between the two, on average, the area of the mRNA sample was 9.6 times greater than that of the dsRNA sample. This observation prompted us to evaluate how each molecule type responded to distinct types of stains to understand the mechanism behind this disproportionate change better.

When the custom, long samples were analyzed using both stain types (Fig. 5.5 and Fig. S5.5, Tables S5.4, S5.5), we again observed that dsRNA fluoresces more with SYTO 61 while mRNA

with RiboRed. Furthermore, for both the long dsRNA and mRNA fragments, the area (Fig. 5.5B and D) showed more statistical difference and higher predictor potential than the height (Fig. S5.5). In addition, Fig. 5.5 also highlights an area variability across different experimental repeats, which were run on different days and/or with chip washes in between experimental repeats. These changes in area stem from changes introduced by the use of different chips and subtle changes to the glass–liquid interface within the microfluidic channels across time. To minimize the effect of these external factors on the peak areas, it is important to run the samples subsequently on the same chip with the different dyes.

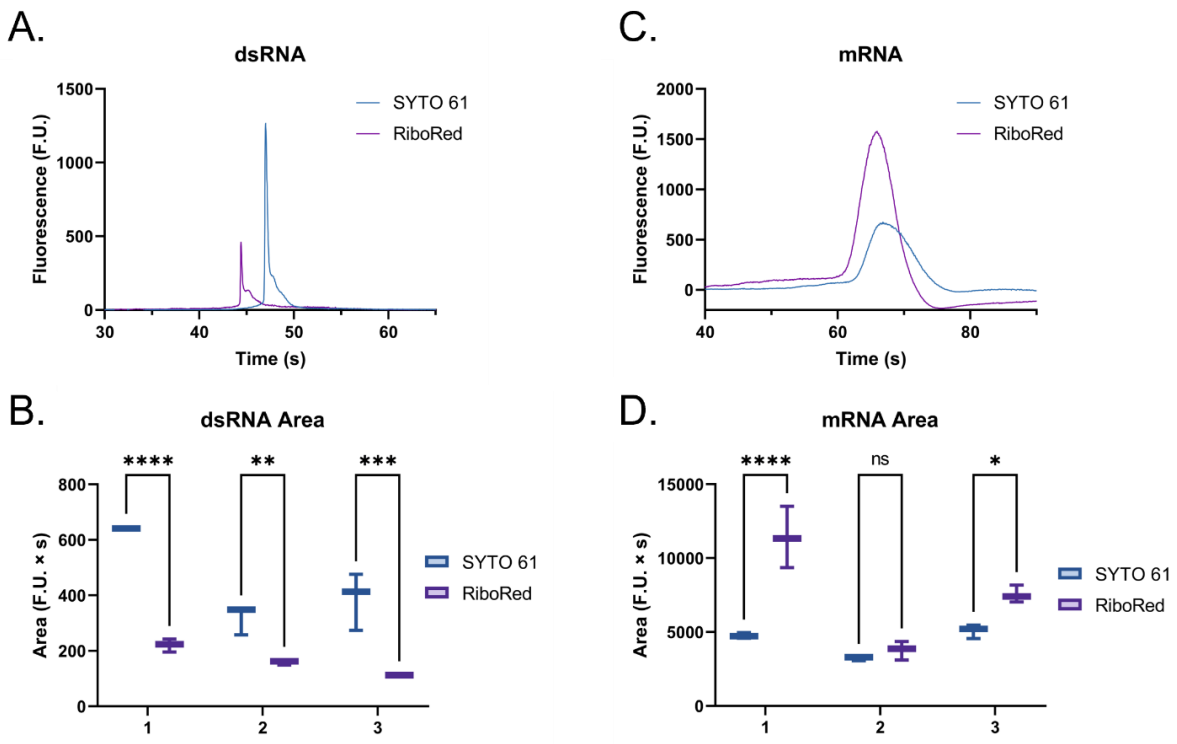

Figure 5.5: Custom dsRNA fragment (A) electropherogram and (B) area, and custom mRNA fragment (C) electropherogram and (D) area using SYTO 61 (blue) and RiboRed (purple). In panels (B) and (D), the numbers on the x-axis represent the number of the experimental repeat, each with two to three instrumental repeats. A slight shift (~5%) is observed in the electropherograms when overlaying both stains, potentially due to differences in stain charge and gel-dye mixture conductivity. However, since the same shift is observed in both molecule types and affects all peaks equally, it is not believed to affect the study. In the comparisons, "*" indicates the significance level in the difference (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$), while "ns" indicates no statistical difference.

Then, to simulate a realistic scenario, we mixed the dsRNA and mRNA samples (Fig. 5.6) at a 1 : 2 ratio and assessed their response in the presence of each fluorophore (Fig. 5.6 and Fig. S5.6, Tables S5.6, S5.7), where the same behavior was observed.

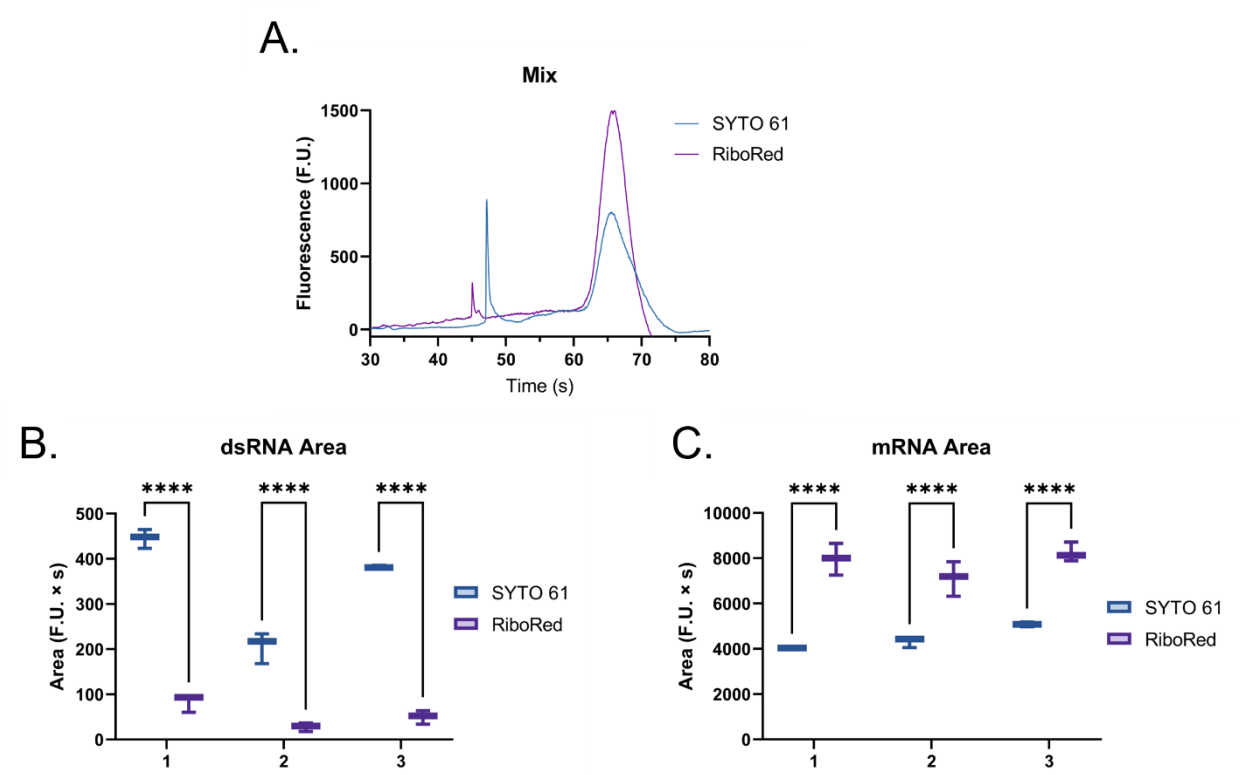

Figure 5.6: (A) Electropherogram of mixed long dsRNA (first peak from left to right with a migration between 45–50 s) and mRNA (second peak from left to right with a migration between 60–70 s) samples where SYTO 61 is represented in blue and RiboRed is represented in purple. (B) dsRNA and (C) mRNA area under the curve. The summarized data includes results from three independent runs, each with two to three repeats. In the comparisons, "*" indicates the significance level in the difference (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$), while "ns" indicates no statistical difference. Once again, the dsRNA was used at a concentration of 2.60 ng µL−1, while the mRNA was used at a concentration of 5.90 ng µL−1 based on triplicate nanodrop readings.

5.4.3 Differentiation Method for Identification and Relative Quantification of dsRNA Contaminants in mRNA Samples

Considering that, at times, there was no statistical difference between the peak heights and even peak areas, we developed a classifier to determine the type of molecule yielding a peak. Based on

the behavior observed when mRNA, ssRNA and dsRNA were exposed to different fluorescent molecules, we hypothesized that the type of molecule could be determined based on the following relationship:

If: $\frac{Peak\ Area_{RNA\ Dye}}{Peak\ Area_{DNA\ Dye}} < 1$,

then the peak-producing molecule is dsRNA.

Or if: $\frac{Peak\ Area_{RNA\ Dye}}{Peak\ Area_{DNA\ Dye}} > 1$,

then the peak-producing molecule is mRNA (or ssRNA).

As hypothesized, the dsRNA peak area stain ratio was <1 while the mRNA area stain ratio was >1 (Table 5.1). In this case, both RNA types met the differentiation criteria on each run. Interestingly, while the area stain ratios yielded by the mRNA peak in both its pure and mixed form were close in value (1.68 for the pure sample and 1.77 for the mixed sample), the area stain ratios yielded by the dsRNA were much more different (0.40 in its pure form and 0.15 when mixed). An added benefit of using ratios to determine the origin of the peak is that it converts the changes in area to relative terms, enabling the comparison of results across different chips and days, which was not possible when comparing the absolute areas.

Table 5.1: Peak area stain ratios (RiboRed/SYTO 61) of each peak run in three independent experiments with two or three repeats. Here, the dsRNA concentration was 2.60 ng μL−1, while the mRNA concentration was 5.90 ng μL−1 based on triplicate nanodrop readings.

	Pure peak area stain ratio		**Mixed peak area stain ratio**	
Run #	**dsRNA**	**mRNA**	**dsRNA**	**mRNA**
1	0.34 ± 0.03	2.39 ± 0.34	0.19 ± 0.05	1.98 ± 0.18
2	0.51 ± 0.11	1.17 ± 0.15	0.14 ± 0.06	1.64 ± 0.11
3	0.34 ± 0.09	1.49 ± 0.06	0.13 ± 0.04	1.66 ± 0.03
Average	**0.40 ± 0.11**	**1.68 ± 0.58**	**0.15 ± 0.05**	**1.77 ± 0.21**

One of the potential causes of this difference is that rather than comparing independent electropherograms to each other, here we are comparing the peak areas obtained from the same electropherograms, potentially decreasing the error of combining the two. Additionally, it must also be noted that, as can be observed in Fig. 5.6A, there is a baseline increase to the left of the mRNA peak, which is not present when dsRNA alone is analyzed, and is likely due to sample degradation, as reported in the literature, which overlaps with the dsRNA. While it would be preferable if the regions did not overlap, the fact that the peak is still differentiable is encouraging. Moreover, even when averaging the three independent runs, there is still a statistical difference between the value yielded between the dsRNA and mRNA samples (Fig. S5.7). While the mechanisms for dynamic staining of SYTO 61 and RiboRed are not precisely known, our data suggests that they exhibit quantifiable selectivity when staining nucleic acids (Fig. S5.8). Some possible explanations for this selectivity include the increased distances in the spaces and grooves of ssRNA and mRNA due to secondary and tertiary structures that cause hairpins, internal loops, and bulges.

RiboRed may have a larger size, making it unable to fit effectively into the grooves or spaces in-between dsRNA. A previous study had reported using a "door-bolt mechanism" that allows near-infrared fluorescent probes to have specificity for RNA.[212] In contrast, SYTO 61 may be too small to fit and label the ssRNA and mRNA molecules effectively. dsRNA has been known to have an A-form helix, having a length increase per base pair of approximately 2.8 Å and a radius of approximately 1.2 nm, causing it to be around 20% shorter and wider than dsDNA, which has a B-form helix in nature.[213] The shorter and wider spacing within dsRNA base pairs and grooves may explain why most fragments in the ssRNA ladder fluoresced brighter than the dsRNA fragments with SYTO 61. If SYTO 61 was created to bind to dsDNA through groove binding or intercalation, it might be too large to bind dsRNA effectively and instead fit the greater spaces in ssRNA better.

5.4.4 Application

The next step in this study was to assess the applicability of the method. First, the linearity of the area to sample concentration was evaluated for both types of molecules using the SYTO 61 and RiboRed stains (Fig. S5.9). Using the areas yielded by each sample from Fig. S5.9 and from Fig. 5.7, the limit of detection (L.O.D.) for this assay was calculated by determining the minimum concentration that could be observed. This was done by dividing the concentration used by the area yielded, which provides the concentration that would yield an area >1. In the case of both types of RNA, the areas produced by the least efficient stain (RiboRed for dsRNA and SYTO61 for mRNA) were used since the method relies on the ability to detect the fragments using both stains. By dividing the concentration of each sample by the peak area it produced, the L.O.D. of the system for dsRNA was estimated to be between 17.7–76.6 pg μL^{-1} while that of mRNA was estimated to be 1.79–2.83 pg μL^{-1}. The difference in L.O.D. is not surprising given that throughout

the experiments, regardless of the stain, both the ssRNA molecules and the mRNA yielded higher areas. In addition, it is important to note that while the areas tend to get greater as the concentrations increase, it was determined that the maximum concentration of mRNA that could be used without clogging or affecting the chip was 12.99 ng μL−1. Assuming we can indeed detect concentrations of dsRNA as low as 17.7–76.6 pg μL^{-1} if we load 12.99 ng μL^{-1} of mRNA, contaminants in percentages as low as 0.14–0.59% relative to the total mRNA concentration should be detectable. Since the actual percentage of dsRNA present in vaccines can vary depending on the construct and production conditions, it is of great importance to have a system with a robust L.O.D. capable of detecting low levels of contaminants.

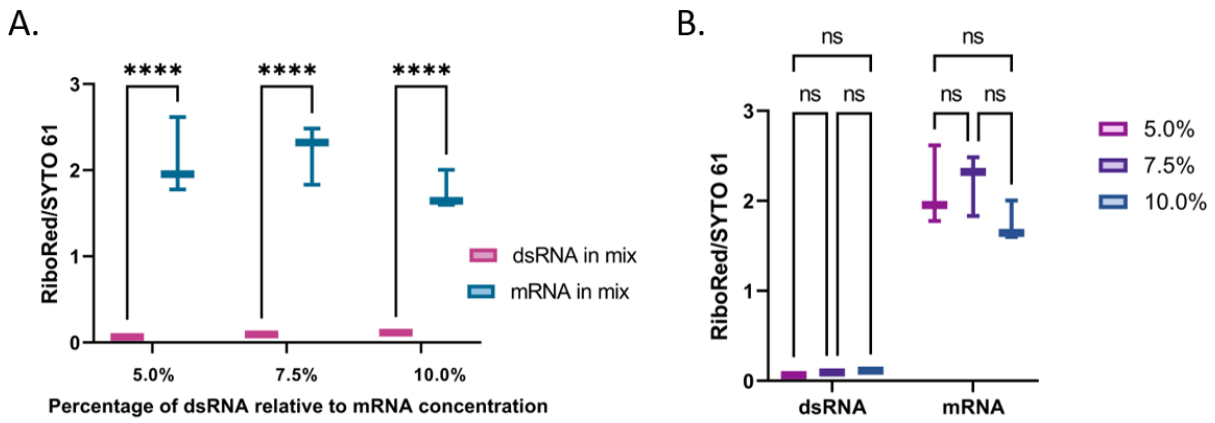

Figure 5.7: Assessment of peak area stain ratios at varying concentrations of dsRNA (5.0%, 7.5%, and 10.0% of the total mRNA concentration, 12.99 ng μL^{-1}). (A) Comparison between the dsRNA (pink) and mRNA (teal) area stain ratios, and (B) comparison among the dsRNA and mRNA area stain ratios, separately, as the relative percentage of dsRNA varied (5.0% pink, 7.5% purple, 10.0% blue). In the comparisons, "*" indicates the significance level in the difference (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$), while "ns" indicates no statistical difference.

The percentage of dsRNA contaminants present in a sample can vary regardless of the mRNA concentration. Therefore, we evaluated whether the validity of the method was dependent on the relative concentration of the contaminant. Irrespective of the dsRNA concentration, there was a difference of equal statistical significance between the dsRNA and the mRNA stain area stain ratios

(Fig. 5.7A and Table S5.8). Similarly, there was no statistical difference between the dsRNA area stain ratios at different concentrations and the mRNA area stain ratio at equal concentrations with varying dsRNA concentrations (Fig. 5.7B and Table S5.8). These results suggest that although the area yielded by each peak will vary based on differences in labeling efficiencies, the method is robust and should yield the expected results regardless of the concentration of dsRNA relative to that of the mRNA. Once the peak has been identified, a standard of known concentration that matches the identified nucleic acid type can be used to estimate the concentration of the sample. We have previously demonstrated the accuracy of using nucleic acid standards (samples of known concentration that undergo the same sample treatment as the other samples) in the context of adeno-associated virus delivery vehicles[109] and are currently developing a similar method for LNP-extracted nucleic acid.

5.5 Conclusions

In this paper, we present a novel method for the identification and relative quantification of dsRNA contaminants in mRNA vaccines. Initially, to assess whether dsRNA and mRNA would exhibit any differences when analyzed in our microfluidic platform, we used a dsRNA ladder to simulate dsRNA contaminants and an ssRNA ladder to simulate mRNA vaccines. After optimizing the sample preparation, gel matrix concentration, and dynamic fluorescent stain concentration, we noted that at the respective optimal dynamic stain concentrations for each fluorophore type, the dsRNA fragments appeared to label more efficiently, 56%, with SYTO 61. In contrast, the ssRNA fragments appeared to label more efficiently, 118%, with RiboRed. Here, we established that the difference in dynamic staining between dsRNA and ssRNA fragments offers a method to identify and quantify the dsRNA percent impurity. This was later validated with long dsRNA and mRNA (4001 bp and nt, respectively) molecules.

As a result, the difference in response between the dsRNA and mRNA molecules to these fluorophores was selected as the basis of our dsRNA contaminant identification and quantitation method. In addition, using the automated LabChip platform, which is compatible with 96- and 364-well plates, makes the platform high-throughput. This characteristic is highly beneficial to assessing dsRNA contaminants within and across mRNA vaccine batches. Although two different gel matrix preparations are required because the samples do not require preparation, the same sample plate is analyzed with both gel matrices limiting the hands-on time to <30 min.

Moreover, the results obtained in this study suggest that concentrations of dsRNA as low as 17.7–76.6 pg μL^{-1} and 1.79–2.83 pg μL^{-1} of mRNA should be detectable using this method. Therefore, when the maximum loading capacity of mRNA (12.99 ng μL^{-1}) is analyzed, dsRNA contaminants present as low as 0.14–0.59% of the total mRNA concentration should be detectable. In addition, since this method allows for the fragment-based detection and characterization of different types of nucleic acid molecules, its analytical potential spans beyond the detection of dsRNA contaminants, as it can also be used to detect partially double-stranded, folded single-stranded, and truncated mRNA fragments. Lastly, other groups have analyzed mRNA samples extracted from LNPs in similar microfluidic chips,[170] which would enable the mRNA composition assessment before and after encapsulation.

5.6 Author Contributions

Conceptualization, A. C. D. P.; data curation, A. C. D. P., N. L., and M. V.; formal analysis, A. C. D. P.; funding acquisition, A. T.; investigation, A. C. D. P, N. L., M. V. and L. B.; methodology, A. C. D. P., L. B., and A. T.; project administration, A. C. D. P., and A. T.; resources, A. T.; supervision, A. C. D. P., and A. T.; validation, A. C. D. P.; visualization, A. C. D. P., and N. L.; writing – original draft, A. C. D. P., N. L., and M. V.; writing – review & editing, A. C. D. P., L. B., and A. T.

5.7 Acknowledgments

This work was partly supported by Revvity's research grant to Brown University. A. T. is a paid scientific advisor/consultant and lecturer from Revvity.

5.8 Supporting Information

Supplementary Materials Information

The ssRNA ladder (NEB catalog #N0364S) contains fragments of 50 nt, 80 nt, 150 nt, 300 nt, 500 nt, and 1,000 nt, while the dsRNA ladder (NEB catalog #N0363S) contains fragments of 21 bp, 30 bp, 50 bp, 80 bp, 150 bp, 300 bp, and 500 bp. It must be noted that the fragment concentration of these samples, in particular the dsRNA ladder, has not been fully characterized by the provider since difficulties in synthesis can lead to significant batch-to-batch differences. The mRNA and dsRNA samples were custom ordered from Genewiz (Genewiz Genomics Headquarters, South Plainfield, NJ) and have lengths of 4,001 nt and 4,001 bp, respectively (provided below). While the sequences were selected to simulate an mRNA vaccine, it must be noted that the product used in this study contains uridines instead of pseudouridines. Unless specified otherwise, all samples were used at room temperature and were not heated prior to analysis. Lastly, all experiments shown include data from three independent runs/experimental repeats (fresh samples, gel-dye, and chip preparation) with two to three repeats/instrumental repeats (same sample gets analyzed multiple times in the same gel-dye and chip preparation), leading to a total of 6-9 data points per data set presented.

Sequence. mRNA and dsRNA Sample Sequence

The custom, long mRNA and dsRNA samples were ordered from Genewiz (Genewiz Genomics Headquarters) and have lengths of 4,001 nt and bp, respectively. The dsRNA is a duplex form of the mRNA sample and as such they share sequences. The sequence for the mRNA and dsRNA 5' strand is provided below. The sequence is believed to have a limited or even negligible effect on the study itself, but it is provided in case it is useful to the reader.

GGGAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAGACCCCGGCGCCGCC
ACCAACCUGACCACCCGGACCCAGCUGCCACCAGCCUACACCAACAGCUUCACCCG
GGGCGUCUACUACCCCGACAAGGUUUCGUGUUCCUGGUGCUGCUGCCCCUGGUGA
GCAGCCAGUGCGUGGUUCCGGAGCAGCGUCCUGCACAGCACCCAGGACCUGUUCC
UGCCCUUCUUCAGCAACGUGACCUGGUUCCACGCCAUCCACGUGAGCGGCACCAA
CGGCACCAAGCGGUUCGACAACCCCGUGCUGCCCUUCAACGACGGCGUGUACUUC
GCCAGCACCGAGAAGAGCAACAUCAUCCGGGGCUGGAUCUUCGGCACCACCCUGG
ACAGCAAGACCCAGAGCCUGCUGAUCGUGAAUAACGCCACCAACGUGGUGAUCAA
GGUGUGCGAGUUCCAGUUCUGCAACGACCCCUUCCUGGGCGUGUACUACCACAAG
AACAACAAGAGCUGGAUGGAGAGCGAGUUCCGGGUGUACAGCAGCGCCAACAAC
UGCACCUUCGAGUACGUGAGCCAGCCCUUCCUGAUGGACCUGGAGGGCAAGCAGG
GCAACUUCAAGAACCUGCGGGAGUUCGUGUUCAAGAACAUCGACGGCUACUUCAA
GAUCUACAGCAAGCACACCCCAAUCAACCUGGUGCGGGAUCUGCCCCAGGGCUUC
UCAGCCCUGGAGCCCCUGGUGGACCUGCCCAUCGGCAUCAACAUCACCCGGUUCC
AGACCCUGCUGGCCCUGCACCGGAGCUACCUGACCCCAGGCGACAGCAGCAGCGG
GUGGACAGCAGGCGCGGCUGCUUACUACGUGGGCUACCUGCAGCCCCGGACCUUC
CUGCUGAAGUACAACGAGAACGGCACCAUCACCGACGCCGUGGACUGCGCCCUGG
ACCCUCUGAGCGAGACCAAGUGCACCCUGAAGAGCUUCACCGUGGAGAAGGGCAU
CUACCAGACCAGCAACUUCCGGGUGCAGCCCACCGAGAGCAUCGUGCGGUUCCCC
AACAUCACCAACCUGUGCCCCUUCGGCGAGGUGUUCAACGCCACCCGGUUCGCCA
GCGUGUACGCCUGGAACCGGAAGCGGAUCAGCAACUGCGUGGCCGACUACAGCGU
GCUGUACAACAGCGCCAGCUUCAGCACCUUCAAGUGCUACGGCGUGAGCCCCACC
AAGCUGAACGACCUGUGCUUCACCAACGUGUACGCCGACAGCUUCGUGAUCCGUG
GCGACGAGGUGCGGCAGAUCGCACCCGGCCAGACAGGCAAGAUCGCCGACUACAA
CUACAAGCUGCCCGACGACUUCACCGGCUGCGUGAUCGCCUGGAACAGCAACAAC
CUCGACAGCAAGGUGGGCGGCAACUACAACUACCUGUACCGGCUGUUCCGGAAGA
GCAACCUGAAGCCCUUCGAGCGGGACAUCAGCACCGAGAUCUACCAAGCCGGCUC
CACCCCUUGCAACGGCGUGGAGGGCUUCAACUGCUACUUCCCUCUGCAGAGCUAC
GGCUUCCAGCCCACCAACGGCGUGGGCUACCAGCCCUACCGGGUGGUGGUGCUGA
GCUUCGAGCUGCUGCACGCCCCAGCCACCGUGUGUGGCCCCAAGAAGAGCACCAA
CCUGGUGAAGAACAAGUGCGUGAACUUCAACUUCAACGGCCUUACCGGCACCGGC
GUGCUGACCGAGAGCAACAAGAAAUUCCUGCCCUUUCAGCAGUUCGGCCGGGACA
UCGCCGACACCACCGACGCUGUGCGGGAUCCCCAGACCCUGGAGAUCCUGGACAU
CACCCCUUGCAGCUUCGGCGGCGUGAGCGUGAUCACCCCAGGCACCAACACCAGC

AACCAGGUGGCCGUGCUGUACCAGGACGUGAACUGCACCGAGGUGCCCGUGGCCA
UCCACGCCGACCAGCUGACACCCACCUGGCGGGUCUACAGCACCGGCAGCAACGU
GUUCCAGACCCGGGCCGGUUGCCUGAUCGGCGCCGAGCACGUGAACAACAGCUAC
GAGUGCGACAUCCCCAUCGGCGCCGGCAUCUGUGCCAGCUACCAGACCCAGACCA
AUUCACCCCGGAGGGCAAGGAGCGUGGCCAGCCAGAGCAUCAUCGCCUACACCAU
GAGCCUGGGCGCCGAGAACAGCGUGGCCUACAGCAACAACAGCAUCGCCAUCCCC
ACCAACUUCACCAUCAGCGUGACCACCGAGAUUCUGCCCGUGAGCAUGACCAAGA
CCAGCGUGGACUGCACCAUGUACAUCUGCGGCGACAGCACCGAGUGCAGCAACCU
GCUGCUGCAGUACGGCAGCUUCUGCACCCAGCUGAACCGGGCCCUGACCGGCAUC
GCCGUGGAGCAGGACAAGAACACCCAGGAGGUGUUCGCCCAGGUGAAGCAGAUCU
ACAAGACCCCUCCCAUCAAGGACUUCGGCGGCUUCAACUUCAGCCAGAUCCUGCC
CGACCCCAGCAAGCCCAGCAAGCGGAGCUUCAUCGAGGACCUGCUGUUCAACAAG
GUGACCCUAGCCGACGCCGGCUUCAUCAAGCAGUACGGCGACUGCCUCGGCGACA
UAGCCGCCCGGGACCUGAUCUGCGCCCAGAAGUUCAACGGCCUGACCGUGCUGCC
UCCCCUGCUGACCGACGAGAUGAUCGCCCAGUACACCAGCGCCCUGUUAGCCGGA
ACCAUCACCAGCGGCUGGACUUUCGGCGCUGGAGCCGCUCUGCAGAUCCCCUUCG
CCAUGCAGAUGGCCUACCGGUUCAACGGCAUCGGCGUGACCCAGAACGUGCUGUA
CGAGAACCAGAAGCUGAUCGCCAACCAGUUCAACAGCGCCAUCGGCAAGAUCCAG
GACAGCCUGAGCAGCACCGCUAGCGCCCUGGGCAAGCUGCAGGACGUGGUGAACC
AGAACGCCCAGGCCCUGAACACCCUGGUGAAGCAGCUGAGCAGCAACUUCGGCGC
CAUCAGCAGCGUGCUGAACGACAUCCUGAGCCGGCUGGACCCUCCCGAGGCCGAG
GUGCAGAUCGACCGGCUGAUCACUGGCCGGCUGCAGAGCCUGCAGACCUACGUGA
CCCAGCAGCUGAUCCGGGCCGCCGAGAUUCGGGCCAGCGCCAACCUGGCCGCCAC
CAAGAUGAGCGAGUGCGUGCUGGGCCAGAGCAAGCGGGUGGACUUCUGCGGCAA
GGGCUACCACCUGAUGAGCUUUCCCCAGAGCGCACCCCACGGAGUGGUGUUCCUG
CACGUGACCUACGUGCCCGCCCAGGAGAAGAACUUCACCACCGCCCCAGCCAUCU
GCCACGACGGCAAGGCCCACUUUCCCCGGGAGGGCGUGUUCGUGAGCAACGGCAC
CCACUGGUUCGUGACCCAGCGGAACUUCUACGAGCCCCAGAUCAUCACCACCGAC
AACACCUUCGUGAGCGGCAACUGCGACGUGGUGAUCGGCAUCGUGAACAACACCG
UGUACGAUCCCCUGCAGCCCGAGCUGGACAGCUUCAAGGAGGAGCUGGACAAGUA
CUUCAAGAAUCACACCAGCCCCGACGUGGACCUGGGCGACAUCAGCGGCAUCAAC
GCCAGCGUGGUGAACAUCCAGAAGGAGAUCGAUCGGCUGAACGAGGUGGCCAAG
AACCUGAACGAGAGCCUGAUCGACCUGCAGGAGCUGGGCAAGUACGAGCAGUACA
UCAAGUGGCCCUGGUACAUCUGGCUGGGCUUCAUCGCCGGCCUGAUCGCCAUCGU
GAUGGUGACCAUCAUGCUGUGCUGCAUGACCAGCUGCUGCAGCUGCCUGAAGGGC
UGUUGCAGCUGCGGCAGCUGCUGCAAGUUCGACGAGGACGACAGCGAGCCCGUGC
UGAAGGGCGUGAAGCUGCACUACACCUGAUAAUAGGCUGGAGCCUCGGUGGCCU
AGCUUCUUGCCCCUUGGGCCUCCCCCCAGCCCCUCCUCCCCUUCCUGCACCCGUAC
CCCCGUGGUCUUUGAAUAAAGUCUGAGUGGGCGGCAAAAAAAAA

Figure S5.1: Assessment of the Effect of Heat During Sample Preparation

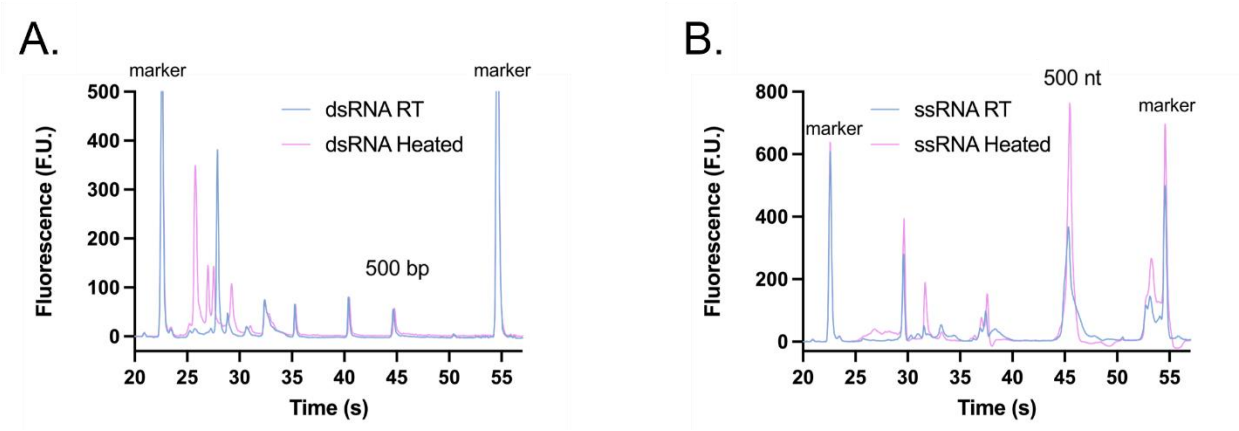

Figure S5.1: Assessment of the effect of heat during dsRNA and ssRNA sample preparation. Generally, RNA molecules that are single-stranded are heated prior to analysis to prevent their tertiary structure (i.e., hairpins) from interfering with the accurate measurement of their mobility. However, upon analyzing both ssRNA and dsRNA (Figure S1) it was observed that while the use of heat resulted in better defined peaks for the ssRNA sample, it also denatured smaller strands of dsRNA causing the fabrication of dsRNA peaks. Acknowledging that numerous techniques exist for the analysis of mRNA molecules, and that the main focus of this study is to characterize the harmful dsRNA contaminant in mRNA vaccines, it was decided that the conditions would be selected to optimize the detection and characterization of dsRNA instead. This denaturing makes the dsRNA more difficult to characterize, since dsRNA can be present in mRNA vaccines in any length equal to or inferior to the mRNA length. Due to dsRNA being the main focus of this work, it was decided to prioritize all lengths of dsRNA over that of ssRNA fragments. Therefore, heating was excluded from the sample preparation. These experiments were conducted using a 6% gel matrix. In addition, while it is difficult to infer from these electropherogram snapshots, the 500 bp dsRNA peak (the last one of the dsRNA ladder) has a migration time of ~45 s while the 500 nt ssRNA peak (the peak after the peak with double intensity in the ladder, which is 300 nt long) has a migration time of ~53 s. Note that the first and last peak in both electropherograms are a lower (left) and an upper (right) marker used to align the electropherograms. Since the upper marker (right) overlaps with some of the larger ssRNA fragments, it was excluded for the study.

Figure S5.2: Optimization of Stain Concentration Based on Peak Response

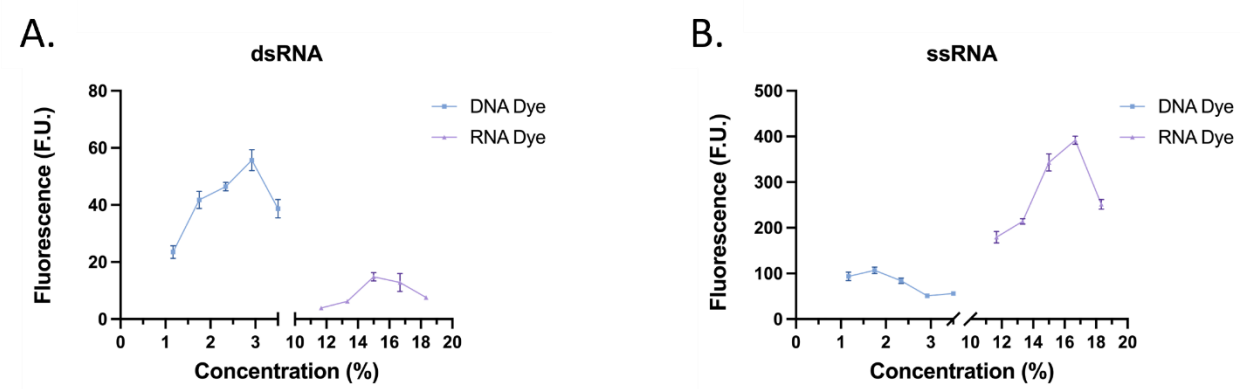

Figure S5.2: Optimization of the ideal dye concentration to detect and characterize both dsRNA and ssRNA fragments based on the response of the 50 bp or nt fragment of each ladder. The fragment size was selected arbitrarily but is believed to be representative of the behavior across each ladder. These experiments were conducted using a 6% gel matrix. When dye concentrations were changed, DMSO was used to keep the gel matrix to dye volume ratio constant for a given dye type. More specifically, the DNA dye (ThermoFisher/Revvity) was diluted to 1.17%, 1.75%, 2.34%, 2.92% and 3.50% dye concentration while the RNA dye (ThermoFisher/Revvity) was diluted to 11.67%, 13.33%, 15.00%, 16.67%, and 18.33% dye concentration. The percentages are all relative concentrations based on the stock vials provided by Revvity of proprietary composition and concentration and are therefore not indicative of the molarity of the solution. To determine the dye concentration of interest (that produced the desirable outcome), we arbitrarily selected a starting concentration for each dye based on provider recommendations, and then probed the neighboring concentrations until a maximum labeling efficiency was achieved.

Figure S5.3: Selection of the Gel Matrix Concentration

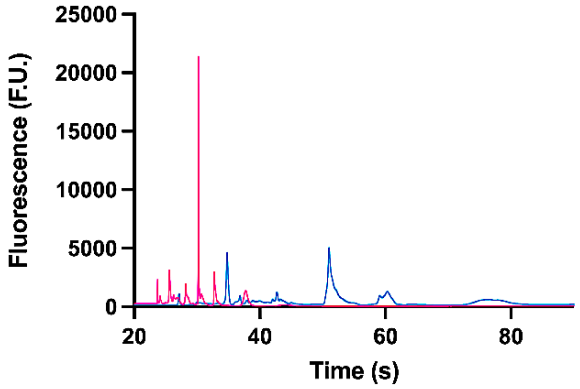

Figure S5.3: The same ssRNA ladder was analyzed first using 6% gel (teal) and then using a 3% gel (pink) matrix mixture. During these experiments it was noted that likely due, or at least in part, to the significantly slower migration using the 6% gel matrix, the ladder showed a significantly greater spread of the peaks at 6% gel than at 3% gel. For this reason, 3% gel was selected as the matrix for the remainder of the experiments. Although not included here, the dsRNA ladder was also assessed with both gel percentages and the results looked similar, although the absence of a larger fragment such as the 1,000 nt present in the ssRNA ladder made the changes in spread less obvious in the dsRNA ladder.

Table S5.1: Comparison of dsRNA and ssRNA Mobility

Table S5.1: Comparison of the migration of different fragments in a dsRNA and a ssRNA ladder within a microfluidic platform. These experiments were conducted using a DNA dye and 3% gel matrix.

Fragment size (bp or nt)	dsRNA migration time (s)	dsRNA mobility ([µm/s]/[V/cm])	ssRNA migration time (s)	ssRNA mobility ([µm/s]/[V/cm])
50	24.5 ± 0.2	1.39 ± 0.01	27.9 ± 1.0	1.38 ± 0.05
80	28.5 ± 0.2	1.35 ± 0.01	*	*
150	30.2 ± 0.3	1.27 ± 0.01	31.47 ± 1.0	1.22 ± 0.04
300	34.1 ± 0.3	1.13 ± 0.01	34.5 ± 0.7	1.11 ± 0.02
500	37.2 ± 0.3	1.03 ± 0.01	37.1 ± 0.7	1.03 ± 0.02

* Note that some fragments were excluded from the analysis due to difficulty distinguishing it from other fragments or from the noise.

Table S5.2: Comparison of dsRNA Peak Response

Table S5.2: Comparison of the dsRNA fragment peak areas and heights using different dye types. These experiments were conducted using a DNA dye at 1.2%, an RNA dye at 15.0%, when applicable, and a 3% gel matrix.

	Area		Height	
Fragment size (bp)	**Area DNA dye (F.U. × s)**	**Area RNA dye (F.U. × s)**	**Height DNA dye (F.U.)**	**Height RNA dye (F.U.)**
21	1458.0 ± 22.8	724.2 ± 29.7	7855.5 ± 114.3	4129.8 ± 102.9
30	345.0 ± 27.5	51.3 ± 1.7	1628.6 ± 91.7	299.3 ± 4.8
150	142.9 ± 7.8	46.7 ± 9.8	804.9 ± 37.9	255.8 ± 26.0
300	226.5 ± 14.9	90.9 ± 19.9	967.8 ± 97.8	284.4 ± 22.1
500	346.6 ± 25.0	196.1 ± 22.8	930.7 ± 50.9	226.2 ± 14.2

Table S5.3: Comparison of ssRNA Peak Response

Table S5.3: Comparison of the ssRNA fragment peak areas and heights using different dye types. These experiments were conducted using a DNA dye at 1.2%, an RNA dye at 15.0%, when applicable, and a 3% gel matrix.

	Area		Height	
Fragment size (nt)	**Area DNA dye (F.U. × s)**	**Area RNA dye (F.U. × s)**	**Height DNA dye (F.U.)**	**Height RNA dye (F.U.)**
50	772.7 ± 124.8	1883.5 ± 12.9	31341 ± 568.7	10287.0 ± 244.9
150	662.2 ± 128.3	1434.8 ± 23.4	1091.3 ± 191.5	2214.7 ± 58.6
300	2157.5 ± 596.9	6566.3 ± 80.1	5423.5 ± 1918.5	11187.7 ± 451.4
500	1565.7 ± 104.7	2296.8 ± 59.5	4845.2 ± 1319.0	5103.6 ± 226.6
1000	1246.9 ± 66.7	1761.1 ± 60.9	1197.8 ± 122.7	1932.8 ± 68.4

Figure S5.4: Comparison of Ladder Peak Height Response using Different Stains

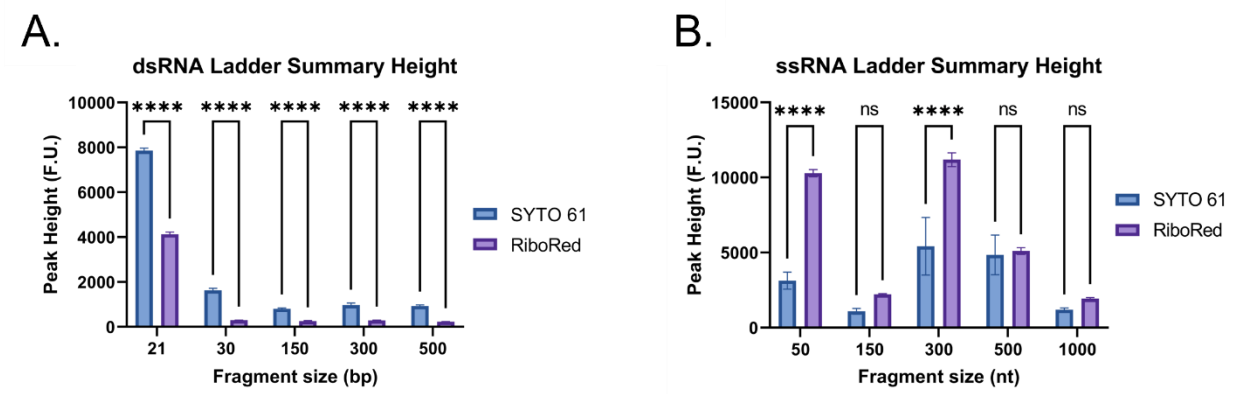

Figure S5.4: Fluorescent staining differences between a dsRNA and an ssRNA ladder using SYTO 61 (blue) and RiboRed (purple). Peak heights produced by (A) the dsRNA ladder and (B) the ssRNA ladder with each stain. In the bar graphs, "*" indicates the level of significance in the difference (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001), while "ns" indicates that there is no statistical difference.

Figure S5.5: Comparison of Pure Long dsRNA and mRNA Peak Height Response using Different Stains

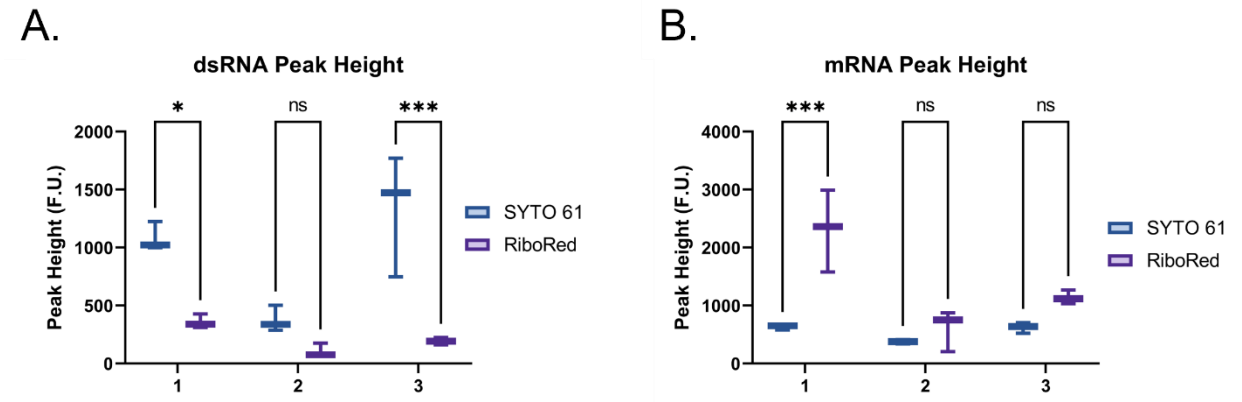

Figure S5.5: Custom pure (A) dsRNA and (B) mRNA fragment peak height using SYTO 61 (blue) and RiboRed (purple). The data includes results from three independent runs, each with two to three repeats. In the comparisons, "*" indicates the level of significance in the difference (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001), while "ns" indicates that there is no statistical difference.

Table S5.4: Comparison of Long dsRNA Fragment Peak Response

Table S5.4: Comparison of a long dsRNA fragment peak areas and heights using different dye types. These experiments were conducted using a DNA dye at 1.2%, an RNA dye at 15.0%, when applicable, and a 3% gel matrix.

	Area		Height	
Run #	**Area DNA dye (F.U. × s)**	**Area RNA dye (F.U. × s)**	**Height DNA dye (F.U.)**	**Height RNA dye (F.U.)**
1	641.3 ± 4.9	220.5 ± 23.5	1082.2 ± 124.2	359.4 ± 61.1
2	319.4 ± 53.8	159.4 ± 9.3	376.0 ± 112.7	107.8 ± 59.1
3	387.3 ± 103.5	112.0 ± 4.0	1330.4 ± 526.3	192.7 ± 45.5
Average	**449.3 ± 158.1**	**170.48 ± 47.8**	**929.5 ± 510.1**	**223.4 ± 127.6**

Table S5.5: Comparison of Long mRNA Fragment Peak Response

Table S5.5: Comparison of a long mRNA fragment peak areas and heights using different dye types. These experiments were conducted using a DNA dye at 1.2%, an RNA dye at 15.0%, when applicable, and a 3% gel matrix.

	Area		Height	
Run #	**Area DNA dye (F.U. × s)**	**Area RNA dye (F.U. × s)**	**Height DNA dye (F.U.)**	**Height RNA dye (F.U.)**
1	4749.1 ± 186.3	11399.7 ± 2077.4	626.3 ± 42.6	2310.0 ± 706.8
2	3220.6 ± 141.7	3777.0 ± 632.6	375.3 ± 36.7	611.4 ± 356.0
3	5082.6 ± 469.0	7536.2 ± 576.6	622.2 ± 93.9	1140.4 ± 119.6
Average	**4350.8 ± 899.0**	**7571.0 ± 3486.8**	**541.3 ± 136.0**	**1354.0 ± 852.5**

Figure S5.6: Comparison of Mixed Long dsRNA and mRNA Peak Height Response using Different Stains

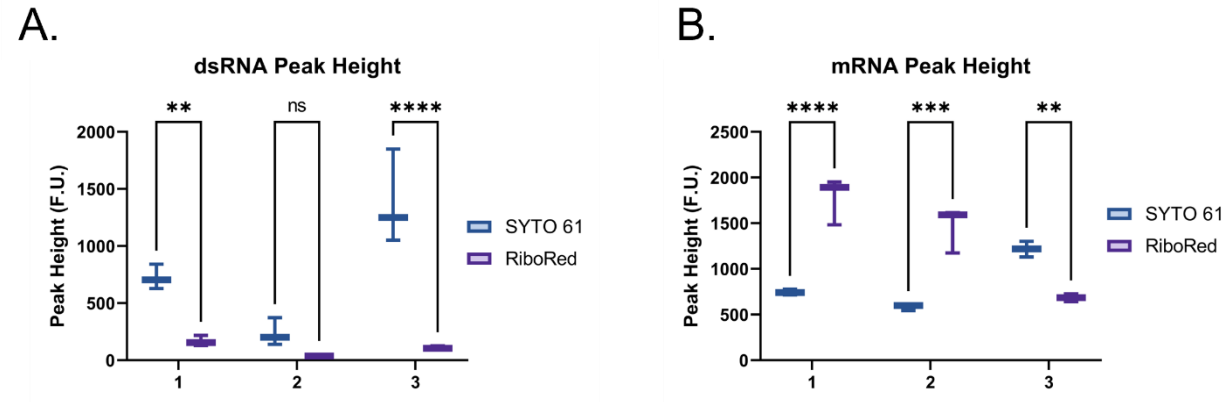

Figure S5.6: Custom mixed (A) dsRNA and (B) mRNA fragment peak height using SYTO 61 (blue) and RiboRed (purple). The data includes results from three independent runs, each with two to three repeats. In the comparisons, "*" indicates the level of significance in the difference (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001), while "ns" indicates that there is no statistical difference.

Table S5.6: Comparison of Long dsRNA Fragment Peak Response When Mixed with mRNA

Table S5.6: Comparison of a long dsRNA fragment peak areas and heights using different dye types when mixed with mRNA. These experiments were conducted using a DNA dye at 1.2%, an RNA dye at 15.0%, when applicable, and a 3% gel matrix. Here, the dsRNA was used at a concentration of 2.60 ng/µL, while the mRNA was used at a concentration of 5.90 ng/µL based on triplicate nanodrop readings.

	Area		**Height**	
Run #	**Area DNA dye (F.U. × s)**	**Area RNA dye (F.U. × s)**	**Height DNA dye (F.U.)**	**Height RNA dye (F.U.)**
1	445.5 ± 20.8	83.6 ± 20.3	724.2 ± 108.5	166.7 ± 46.2
2	206.4 ± 34.2	27.9 ± 9.6	237.2 ± 120.9	38.6 ± 11.2
3	381.0 ± 4.5	50.1 ± 15.1	1383.1 ± 415.8	110.0 ± 14.3
Average	**344.3 ± 109.0**	**53.9 ± 27.8**	**781.5 ± 545.8**	**105.1 ± 495.0**

Table S5.7: Comparison of Long mRNA Fragment Peak Response When Mixed with dsRNA

Table S5.6: Comparison of a long mRNA fragment peak areas and heights using different dye types when mixed with dsRNA. These experiments were conducted using a DNA dye at 1.2%, an RNA dye at 15.0%, when applicable, and a 3% gel matrix. Here, the dsRNA was used at a concentration of 2.60 ng/µL, while the mRNA was used at a concentration of 5.90 ng/µL based on triplicate nanodrop readings.

	Area		**Height**	
Run #	**Area DNA dye (F.U. × s)**	**Area RNA dye (F.U. × s)**	**Height DNA dye (F.U.)**	**Height RNA dye (F.U.)**
1	4034.8 ± 37.6	7971.1 ± 700.0	743.6 ± 30.7	1775.7 ± 256.5
2	4328.0 ± 237.2	7120.1 ± 765.7	583.8 ± 35.7	1460.1 ± 248.2
3	5082.8 ± 144.5	8244.2 ± 425.7	1216.6 ± 85.7	684.4 ± 59.8
Average	**4405.7 ± 460.4**	**7778.5 ± 756.4**	**848.0 ± 289.2**	**1384.6 ± 495.0**

Figure S5.7: RiboRed and SYTO61 Peak Area Ratio for dsRNA and mRNA

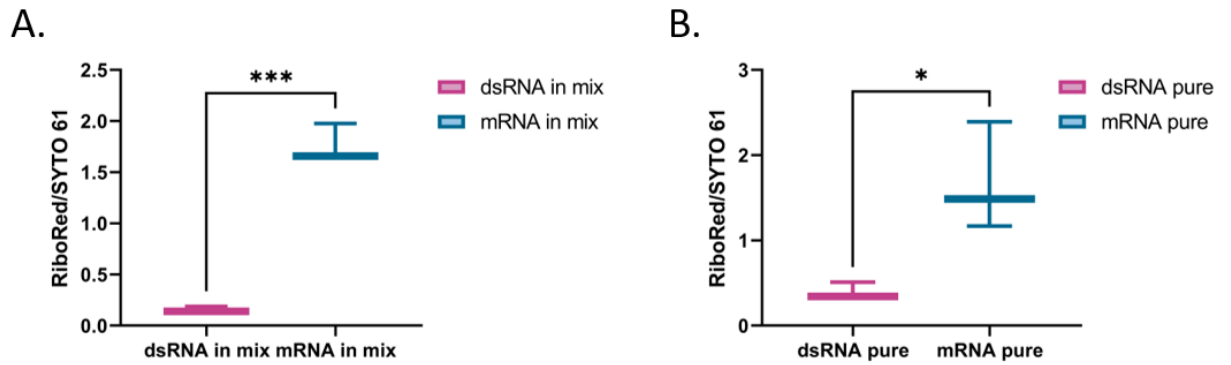

Figure S5.7: Peak area ratios between RiboRed and SYTO 61 for dsRNA (pink) and mRNA (teal) when analyzed (A) mixed and (B) individually. These results include data from three independent runs, each with two to three repeats. Here, it is important to note that the difference in response is exacerbated when comparing the pure to the in-mix dsRNA and mRNA responses. While the exact mechanism behind this response remains unknown, we believe a key factor in this difference is that when comparing the areas yielded by the "pure" RNA molecules, four different electropherograms have to be combined and compared to each other (one per stain type, two per molecule) since the samples are run separately whereas the "in mix" samples only require the combination of two. This can significantly decrease the error associated to the combination of electropherograms since each dsRNA and mRNA peak for each stain type is now being extracted from the same run, which means that the conditions (i.e., conductivity) in which each peak was produced is completely matched across molecule types. In the comparisons, "*" indicates the level of significance in the difference (*$p<0.05$, **$p<0.01$, ***$p<0.001$, ****$p<0.0001$), while "ns" indicates that there is no statistical difference.

Figure S5.8: Interaction between RiboRed and SYTO61 with dsRNA and mRNA

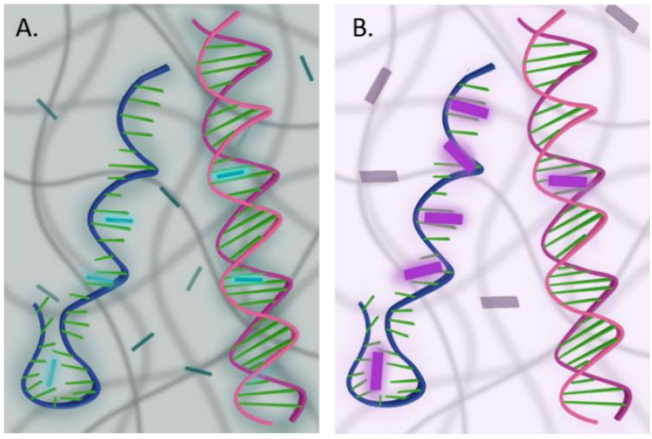

Figure S5.8: Interaction between mRNA (blue) and dsRNA (pink) with (A) SYTO 61 and (B) RiboRed. It is important to note that while dsRNA will label better with SYTO 61 than with RiboRed and the opposite for mRNA, our results suggest that despite this trend, mRNA will still label better than dsRNA with both stains. We hoped to highlight this by showing dsRNA bound to less fluorophores than mRNA even in the optimal labeling conditions for dsRNA.

Figure S5.9: Linearity in dsRNA and mRNA Response

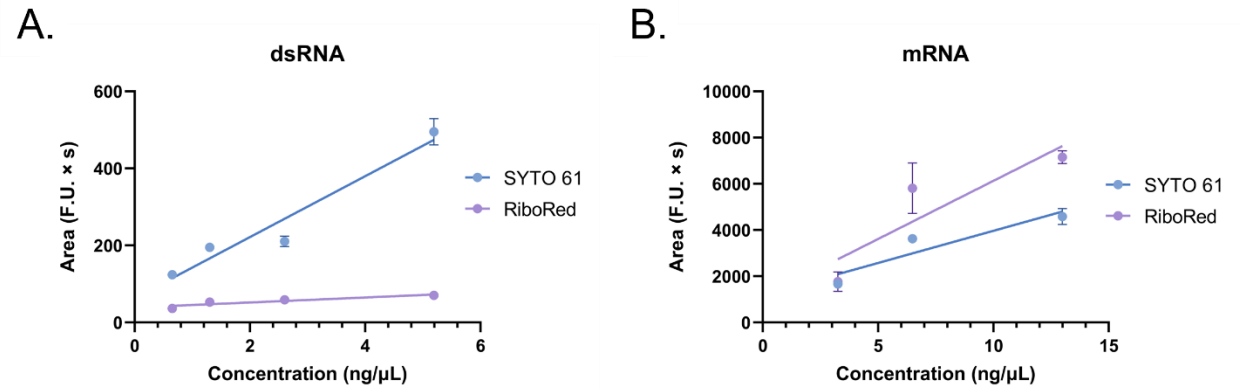

Figure S5.9: Assessment of the linearity in the response of (A) dsRNA and (B) mRNA as their concentrations increased using SYTO 61 (blue) and RiboRed (purple). (A) The line of best fit for the dsRNA sample using SYTO 61 was y=79.05x+63.87 with an R^2=0.93 while that of RiboRed was y=6.49x+39.21 with an R^2=0.75. (B) The line of best fit for the mRNA sample using SYTO 61 was y=278.2x+1189 with an R^2=0.84 while that of RiboRed was y=503.0x+1103 with an R^2=0.75. While all samples showed a roughly linear relation between area and concentration, it was interesting to note that for both dsRNA and mRNA, RiboRed yielded the lowest R^2 value of 0.75 for both molecules, while SYTO 61 yielded the highest, with values of 0.93 and 0.84, respectively. Another interesting trend that was observed during these experiments is that for both molecule types, the slope of the line of best fit was highest with the most efficient stain, SYTO 61 in the case of dsRNA and RiboRed in the case of mRNA. Based on the equations for the lines of best fit yielded for each molecule by both fluorescent stains (Figure 7), in the case of dsRNA the lines would converge when the dsRNA concentration is -0.34 ng/µL while the mRNA lines would converge when the mRNA concentration is 0.38 ng/µL.

Table S5.8: Assessment of dsRNA and mRNA Peak Area Ratios at Varying Concentrations of dsRNA

Table S5.8: Assessment of dsRNA and mRNA peak area ratios (RNA dye/DNA dye) at varying concentrations of dsRNA so that it represented 5.0%, 7.5% and 10.0% of the total mRNA concentration.

dsRNA percentage (%)	dsRNA concentration (ng/µL)	dsRNA peak area ratio	mRNA concentration (ng/µL)	mRNA peak area ratio
5.0	0.65	0.07 ± 0.02	12.99	2.11 ± 0.44
7.5	0.97	0.10 ± 0.00	12.99	2.21 ± 0.34
10.0	1.30	0.11 ± 0.01	12.99	1.75 ± 0.22
Average	**-**	**0.09 ± 0.02**	**-**	**2.03 ± 0.37**

Chapter 6: Enzymatic Isolation and Microfluidic Electrophoresis Characterization of Residual dsRNA Impurities in mRNA Vaccines and Therapeutics

---

Chapter 6 has been submitted as an original research article.

Coll De Peña, A.,[1] Vaduva, M.,[2] Li, N.S.,[1] Shah, S.,[3] Ben Frej, M.,[3] Tripathi, A.,[1]
Enzymatic Isolation and Microfluidic Electrophoresis Characterization of Residual dsRNA Impurities in mRNA Vaccines and Therapeutics, **in review**.

[1]Center for Biomedical Engineering, School of Engineering, Brown University, Providence, RI, USA

[2]Department of Molecular Biology, Cell Biology, and Biochemistry, Division of Biology and Medicine, Brown University, Providence, RI, USA

[3]Applied Genomics, Revvity, Hopkinton, MA, USA

6.1 Abstract

The versatility, rapid development, and ease of production scalability of mRNA therapeutics have placed them at the forefront of biopharmaceutical research. However, despite their vast potential to treat diseases, their novelty comes with unsolved analytical challenges. A key challenge in ensuring sample purity has been monitoring residual, immunostimulatory dsRNA impurities generated during the in-vitro transcription of mRNA. Here, we present a method that combines an enzyme, S1 nuclease, to identify and isolate dsRNA from an mRNA sample with a microfluidic electrophoresis analytical platform to characterize the impurity. After the method was developed and optimized, it was tested with clinically relevant, pseudouridine-modified 700 and 1,800 bp dsRNA and 818-4,451 nt mRNA samples. While the treatment impacted the magnitude of the fluorescent signal used to analyze the samples due to the interference of the buffer with the labeling of the sample, this signal loss was mitigated by 8.8 × via treatment optimization. In addition, despite the mRNA concentration being up to 400 × greater than that of the dsRNA, under every condition, there was a complete disappearance of the main mRNA peak. While the mRNA peak was digested, the dsRNA fragments remained physically unaffected by the treatment, with no change to their migration time. Using these samples, we detected 0.25% dsRNA impurities in mRNA samples using 15 μL with an analytical runtime of 1 min/sample after digestion and were able to predict their size within 8% of the expected length. The short runtime, sample consumption, and high throughput compatibility make it suitable to support the purity assessment of mRNA during purification and downstream.

6.2 Introduction

During the COVID-19 pandemic, the importance of mRNA vaccines was highlighted due to their low production cost, fast manufacturing, and effectiveness, among other advantages.[3,18,214–217] As the first vaccine of its kind to be approved by the FDA, it has accelerated the research interest in mRNA-based vaccines and therapies, showing potential in treating genetic, infectious, cancer, and other diseases and preventing them.[18,217–219] Despite the immense potential of this technology to combat diseases, manufacturing challenges remain that allow for residual impurities in the final product.[2,220,221] Numerous filtration and chromatographic purification steps are integrated into the manufacturing of mRNA to remove the numerous nucleic acid by-products and impurities generated during the process. However, double-stranded RNA (dsRNA), which can be a by-product of the in-vitro transcription (IVT) of mRNA, is a notable challenge for manufacturers, as it behaves similarly to mRNA in current quality control assays and is notoriously difficult to remove during purification.[222–224] To address this, improvements have been made in terms of enzyme technology and purification methods,[225] however, residual dsRNA at trace concentrations can remain in the final product.[226] dsRNA is a molecule of interest due to its ability to significantly increase the immune response to mRNA vaccines,[227] leading to both local and systemic inflammation in treated individuals.[228,229] Further, due to the strong immune response to dsRNA,[230] it has also been found to decrease the effectiveness of the treatment due to the inhibition of translation.[226] The current analytical standard for dsRNA monitoring is dot blot, which is capable of detecting low concentrations of dsRNA contaminants in mRNA samples. Despite its widespread use, dot blot only provides information regarding the concentration and not the length of the dsRNA impurities. Given the potential implications for the efficacy of the therapeutics and the

length-dependent adverse effects dsRNA can cause,[230] it is essential to monitor both the concentration and the length of these impurities.

Microfluidic electrophoresis has been used to analyze mRNA molecules[170,231–234] and can detect dsRNA molecules.[173] Its rapid turnaround time, high throughput compatibility, and low sample consumption[6,51] make it particularly suited to support the development of biopharmaceutical products. However, in addition to the main mRNA peak and the residual dsRNA of interest, other by-products, such as truncated or degraded mRNA fragments, may still be present,[222] making it so that mobility differences alone between mRNA and dsRNA cannot be used for their characterization.[173] Therefore, it is crucial to integrate a method to identify the peak-producing molecules before assessing their size and concentration.

We previously presented an identification method that differentiated dsRNA from mRNA via their dynamic staining response.[173] While it presented an interesting identification method, in practice, the significantly larger mRNA peak tended to mask the significantly smaller dsRNA peak, and the need to use two different gel dyes caused minor shifts in the electropherogram that made it difficult to match the peaks. Instead, this study introduces an analytical approach that utilizes an enzyme, S1 nuclease, to digest all single-stranded fragments in the sample, enabling the characterization of the remaining dsRNA products (Figure 6.1). Post digestion, the samples are analyzed using a custom microfluidic method we developed that enables the detection of low concentrations of dsRNA and mRNA but using a single gel dye mixture.[173] By integrating the isolating effect of S1 nuclease digestion into a microfluidic detection regiment, this method provides a rapid, reproducible, and efficient means for analyzing dsRNA impurities in mRNA formulations.

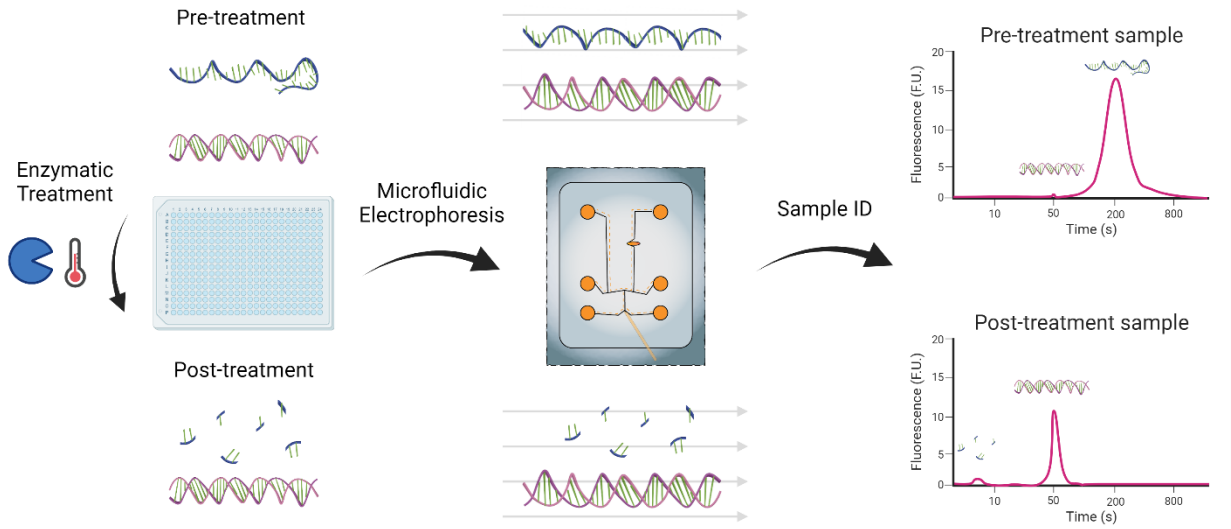

Figure 6.1: Enzymatic identification and microfluidic electrophoresis characterization workflow of dsRNA contaminants and impurities in mRNA samples. Figure created using BioRender.com and Adobe Illustrator.

This study aims to assess the efficacy and specificity of the S1 nuclease-assisted microfluidic approach for detecting dsRNA impurities in mRNA formulations. While the manufacturer-recommended S1 nuclease treatment showed great initial promise, it was initially developed for single-stranded DNA digestion (i.e., overhangs) and had to be optimized to be more fit for purpose. The primary considerations during its optimization were maximizing its compatibility with the microfluidic analytical platform by mitigating its impact on the dsRNA signal and adjusting the reaction buffer to enzyme ratios to ensure the method was robust and would perform well against different lengths and concentrations of mRNA. The optimization of the nuclease treatment enabled an 8.8-fold increase in the dsRNA limit of detection post-digestion from 2.83 ng/µL using the standard protocol to 0.32 ng/µL after optimization. Capable of digesting mRNA concentrations of at least 200 ng/µL, this suggests the method has a theoretical dsRNA contaminant sensitivity of 0.16% of the total mRNA, and the lowest percentage tested was 0.25%, which yielded reproducible results. However, the actual percentage is likely even lower, as the maximum loading of the system was not tested beyond 200 ng/µL, based partly on sample availability.

6.3 Materials and Methods

6.3.1 Materials

The mRNA and dsRNA samples used in this study were pseudouridine chemically modified since this chemical modification has been shown to increase the stability of the molecule and is of great clinical relevance.[235–237] The pseudouridine chemically modified mRNA (818, 1,198, 1,913, 3,406, and 4,451 nt) and dsRNA (700 and 1800 bp) fragments were purchased from CATUG Biotechnology (CatPure™; Cambridge, MA) at respective concentrations of 2,000 ng/µL and 200 ng/µL. The chemically modified samples used in this study are clinically relevant as they were synthesized following the same method (in vitro transcription) and chemical modification (pseudouridine) as the Moderna Therapeutics and Pfizer-BioNTech COVID-19 vaccines.[236] The dsRNA ladder used to assess the length of the impurities was composed of a dsRNA ladder 21-500 bp (New England Biolabs, Ipswich, MA) and a 4,001 bp custom synthesized dsRNA construct (Azenta, Burlington, MA). Before use, the samples were diluted to the specified concentrations in 1X TE buffer, pH 8.0. The S1 nuclease (100 U/µL) and its 5 × reaction buffer were purchased from ThermoFisher Scientific (Franklin, MA) for the enzymatic treatment. They were diluted in nuclease-free water to the specified concentration. S1 nuclease was selected for this study because it is expected to degrade single-stranded nucleic acids (ThermoFisher Scientific), potentially enabling the differentiation between mRNA and dsRNA. For the microfluidic analysis, a custom glass nucleic acid microfluidic chip with a metal sipper and custom chip reagents compatible with the LabChip GXII Touch platform were obtained from Revvity (Waltham, MA).

6.3.2 Enzymatic Treatment

Per the manufacturer, ThermoFisher Scientific, the standard operating procedure (SOP) treatment protocol for the use of S1 Nuclease recommends the use of 10 U (0.1 µL) of enzyme for every 1

μg of nucleic acid and 6 μL of 5 × reaction buffer per 10 U of enzyme, in a total volume of 30 μL. Once mixed, it is incubated for 30 minutes at 25°C, then 2 μL of 0.5 M EDTA are added, and the mixture is heated at 70°C for 10 minutes for enzyme inactivation. Due to the lower concentration at which mRNA samples tend to be available, we adjusted the initial protocol to a total volume of 15 μL and 0.20 μg of nucleic acid (0.150 μg of mRNA and 0.075 μg of dsRNA). As such, the SOP became 2.25 U of the enzyme (2.25 μL of a 1:100 dilution) and 1.35 μL of 5 × reaction buffer for 10 ng/μL of mRNA and 5 ng/μL of dsRNA, mixed with nuclease-free water to a final volume of 15 μL, and then incubated at 25°C for 30 min. When specified, 1 μL of 0.5 M EDTA was added before the heat inactivation of the sample at 70°C for 10 min.

During the study, adjustments to the treatment components will be reported as changes relative to the latter protocol described above. For instance, if 5 U of the enzyme is used instead of the standard 2.25 U, it will be reported as 2 × enzyme. Regardless of the adjustments made to the components, the treatment volume was kept constant across the study. In the case of changes to the nucleic acid concentration, it will refer solely to changes in the mRNA concentration. In contrast, the dsRNA concentration will be kept constant across the individual experiments unless specified for the limit of detection (LOD) assessment.

6.3.3 Microfluidic Measurements and Analysis

Once the samples were loaded onto a 96- or 384-well plate, which could be done before or after digestion, the plate and the microfluidic chip were transferred onto a LabChip GXII Touch platform (Revvity). The pressure, electric fields, and robotic motion to load and analyze the sample onto the chip were controlled using the LabChip platform. We describe this system, the setup, and the electrokinetic script used here in a previous study.[173]

The electropherograms yielded by the LabChip GXII Touch platform were analyzed using the LabChip GX Reviewer software, version 5.11 (Revvity). The data extracted from the LabChip GX Reviewer software was then input into GraphPad Prism 10.1 for statistical analysis and visualization. The statistical significances reported through the study were obtained from a Tukey post hoc test with a confidence interval of 95%, where *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001. In addition, JMP Pro 17 was used for the factorial design and analysis described in the study, which helped determine the main effects and all two-factor interactions of the enzyme, buffer, and mRNA concentrations using three values for each.

To determine the sensitivity of the method, the limit of detection, LOD, of the platform was calculated using the peak area as follows:

$$LOD_{\ area} = \frac{concentration}{peak\ area} \tag{6.1}$$

and via the peak height using the following relation:

$$LOD_{height} = \frac{concentration}{\left(\frac{peak\ height}{3 \times noise}\right)} \tag{6.2}$$

where 3 was used as the desired signal-to-noise ratio, and the noise magnitude of the individual electropherograms was input as the noise.

6.4 Results and Discussion

6.4.1 S1 Nuclease Treatment Baseline

mRNA and dsRNA, particularly longer fragments, have significantly different electrophoretic mobilities.[173] Therefore, to characterize contaminant or residual dsRNA in mRNA samples, it is essential to first identify the peak-producing molecule before using a ladder to determine its length. S1 nuclease is an enzyme that degrades single-stranded nucleic acids and presents a promising

method for identifying dsRNA fragments in mRNA samples prior to microfluidic characterization. To assess the potential of the method, we compared an untreated sample to a sample treated following the standard operating procedure (SOP), which contains EDTA (Figure 6.1a,b). While EDTA is commonly used due to its ability to chelate magnesium ions required by nuclease enzymes, protecting nucleic acids from further degradation, in excess, it can affect the performance of microfluidic analyses. Therefore, we also investigated the need for EDTA in the treatment (Figure 6.1c). When the untreated sample was compared to the treated samples, a complete disappearance, or digestion, of the mRNA peak (~ 37.7 s), but retention, albeit at a significantly lower magnitude, of the dsRNA peak (~ 31.4 s), as highlighted in Figure 1. Interestingly, EDTA did not have a measurable impact on the dsRNA peak area or migration time, and neither did the treatment, regardless of the use of EDTA, compared to the untreated sample. Since the goal is to develop a streamlined and robust protocol for detecting residual dsRNA in mRNA samples, and EDTA did not appear to impact the digestion performance, it was removed from the workflow.

Ensuring that dsRNA remains intact, confirmed by the lack of change in migration time (Figure 6.1f), is crucial for the robustness of the method as it will allow correct length determination downstream. However, while the S1 nuclease treatment continues to show great promise, there was a significant decrease of over two orders of magnitude in the dsRNA peak area, going from 249.8 to an average of 2.0 F.U. × s from untreated to treated. This suggests that the limit of detection (LOD) of the method, at this stage, was 2.83 ng/μL, calculated using the average yielded by Equations 1 and 2. Considering the target is to be able to detect ~ 1% dsRNA by concentration in mRNA samples, and these samples can be present as smears, this would require the loading of over 300 ng/μL of the mRNA sample. Therefore, the next goal of the study is to mitigate the signal decrease of dsRNA caused by the treatment to increase the sensitivity of the method.

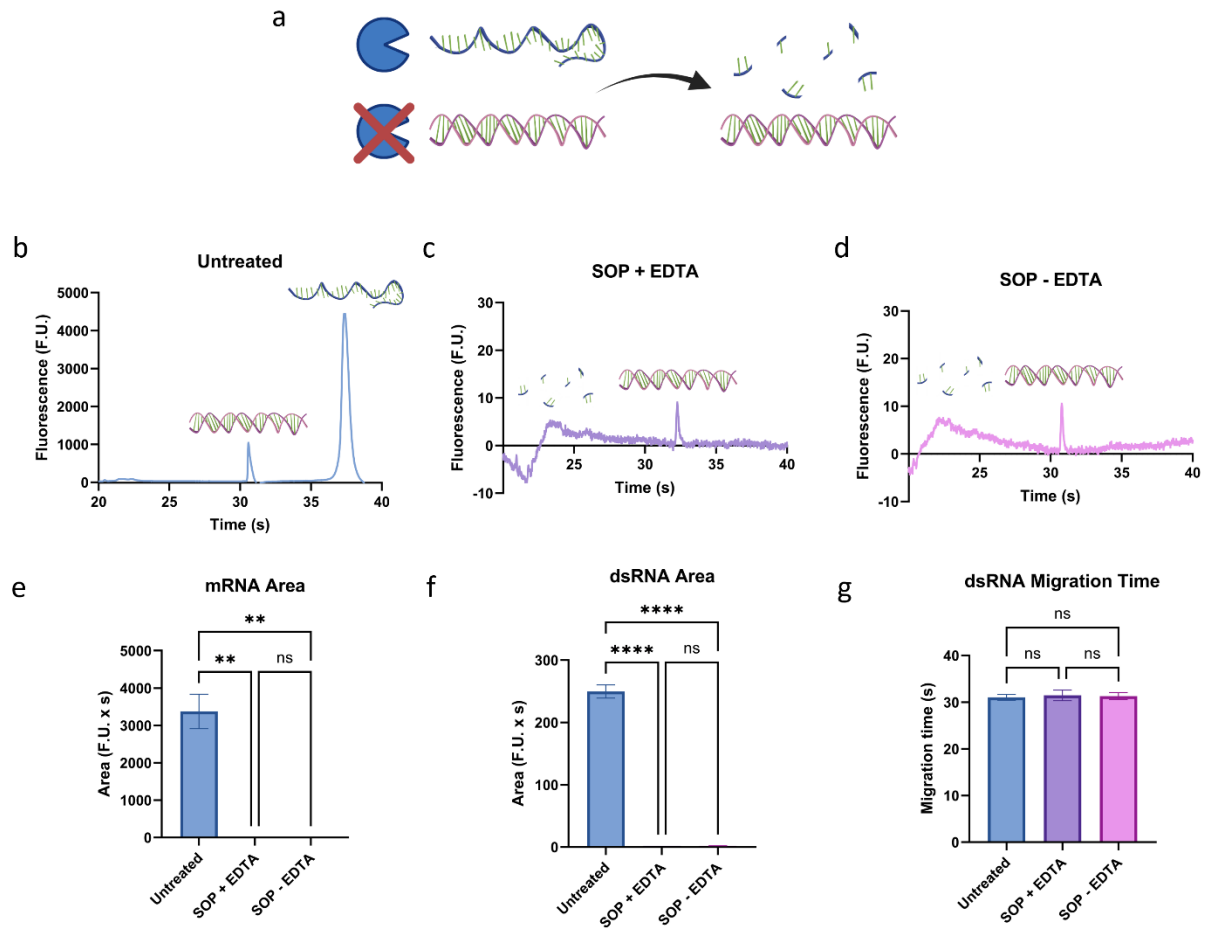

Figure 6.2: Assessment of the impact of the (a) S1 nuclease treatment on mRNA and dsRNA. Electropherograms of 5 ng/µL of a 700 bp dsRNA fragment and 10 ng/µL of a 1,198 nt mRNA fragment (b) untreated, (c) treated following the SOP, which contains EDTA, and (d) treated following the SOP without EDTA. When present, the dsRNA peak has an average migration time of 31.4 s while the mRNA peak has an average migration time of 37.7 s; between 20 and 25 s, a small peak can usually be observed representing the lower marker. Summarized (e) mRNA peak area, (f) dsRNA peak area, and (g) dsRNA peak migration time under the three conditions. "*" indicates the significance level in the difference (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$), while "ns" indicates no statistical difference. This figure was created using BioRender.com and GraphPad Prism.

6.4.2 Optimization of the Buffer to Enzyme Ratio of the Digestion

The relationship between buffer, enzyme, and mRNA concentrations was evaluated to increase the sensitivity of the method. First, some initial experiments not included in this study were conducted

to narrow down the optimal concentrations of each parameter. During these experiments, it was observed that these parameters can interact with each other, so we designed a factorial experiment to test three levels of each of these parameters. We tested buffer concentrations of 1/16, 1/8, and 1/4 × SOP concentration, enzyme concentrations of 2, 10, and 50 × SOP concentration, and mRNA concentrations of 1/2, 1, and 2 the initial concentration of 10 ng/µL (Table S6.1). As can be observed in Figure 6.1, and as we confirmed during initial tests, under different conditions, a smear can appear next to the lower marker. While no experiments were conducted to characterize the nature of the smear further, it is believed it is primarily composed of fragments of mRNA that were not fully digested.

Therefore, to determine the performance of each condition, the dsRNA peak area and height and the undesired peak area were analyzed and compared (Figure S6.1), and the most promising (high dsRNA peak area and height and low undesired peak area) were chosen for further investigation, as highlighted in Figure 6.2a,b. A factorial analysis (Figure S6.2) was used to determine the optimal formulation to maximize the dsRNA peak and minimize the undesired peak. Based on these parameters, the optimal condition was determined to be condition 13, with a buffer concentration of 1/16 × SOP, an enzyme concentration of 10 × SOP, and 1 × the mRNA concentration (Figure 6.2c). Overall, it was observed that inadequate buffer to enzyme concentrations for a given mRNA concentration can lead to insufficient digestion of the mRNA. Considering that using this condition, the LOD is ~ 0.26 ng/µL, which would require a concentration of over 30 ng/µL of mRNA to detect the impurities effectively, it was decided, based on a similar performance, to test further condition 14 (Figure 6.2d), which only differed in buffer concentration (1/8 instead of 1/16 ×) from condition 13, against different mRNA concentrations to ensure the robustness of the method.

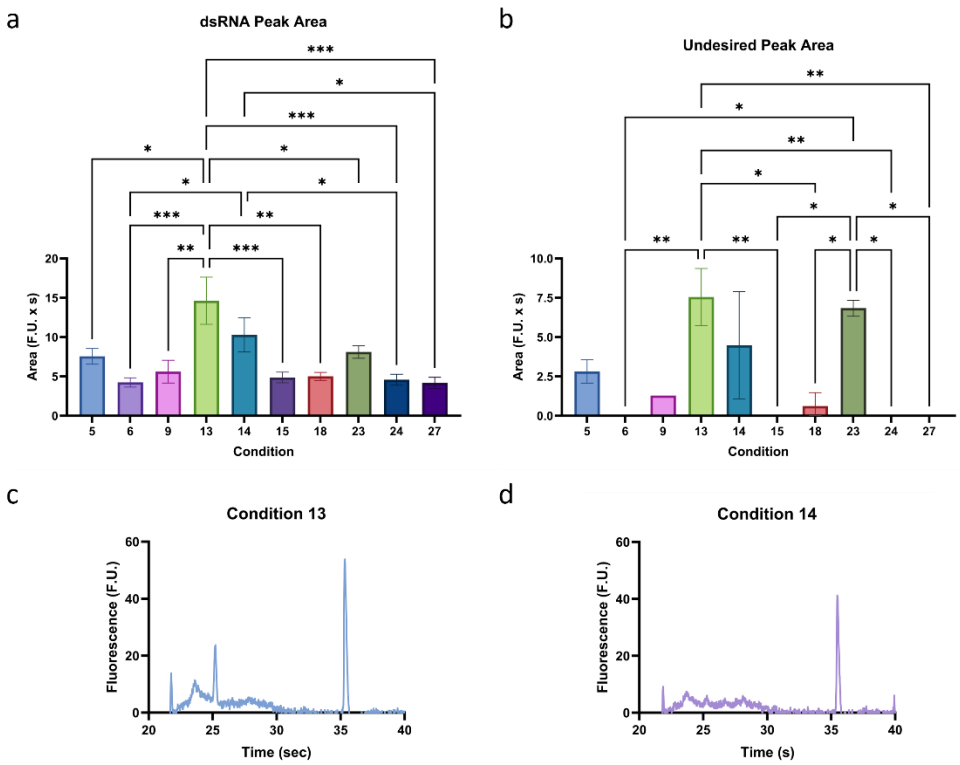

Figure 6.3: Assessment of the optimal buffer to enzyme and mRNA concentrations (Table S1). Summarized (a) dsRNA peak area and (b) undesired peak area of the top performing conditions. Electropherograms of conditions (c) 13 and (d) 14, which were selected for further analysis. "*" indicates the significance level in the difference (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$), while "ns" indicates no statistical difference. This figure was created using GraphPad Prism.

Since condition 13 (1/16 × buffer, 10 × enzyme) was the preferred condition based on the factorial analysis and condition 14 (1/8 × buffer, 10 × enzyme) was semi-qualitatively our preferred condition based on overall peak magnitudes and shapes, we conducted an experiment where mRNA concentrations of 10, 100, and 200 ng/µL were used to test the robustness of each method. We found that the undesired peak area is positively correlated with mRNA concentration for both conditions, and the dsRNA peak area is independent of changes in mRNA concentration (Figure 6.3a,b). Furthermore, the distance between the undesirable and dsRNA peaks in both conditions was sufficient to avoid peak overlap, highlighting the robustness of the method. As predicted, condition 13 had significantly higher undesirable peak areas than condition 14 at higher

concentrations (Figure 6.3c-f). Despite this, the dsRNA peak area was similar for the two conditions, which means that condition 14 has a higher dsRNA peak area to undesirable peak area ratio. As a result, Condition 14 was chosen as the optimal treatment formulation.

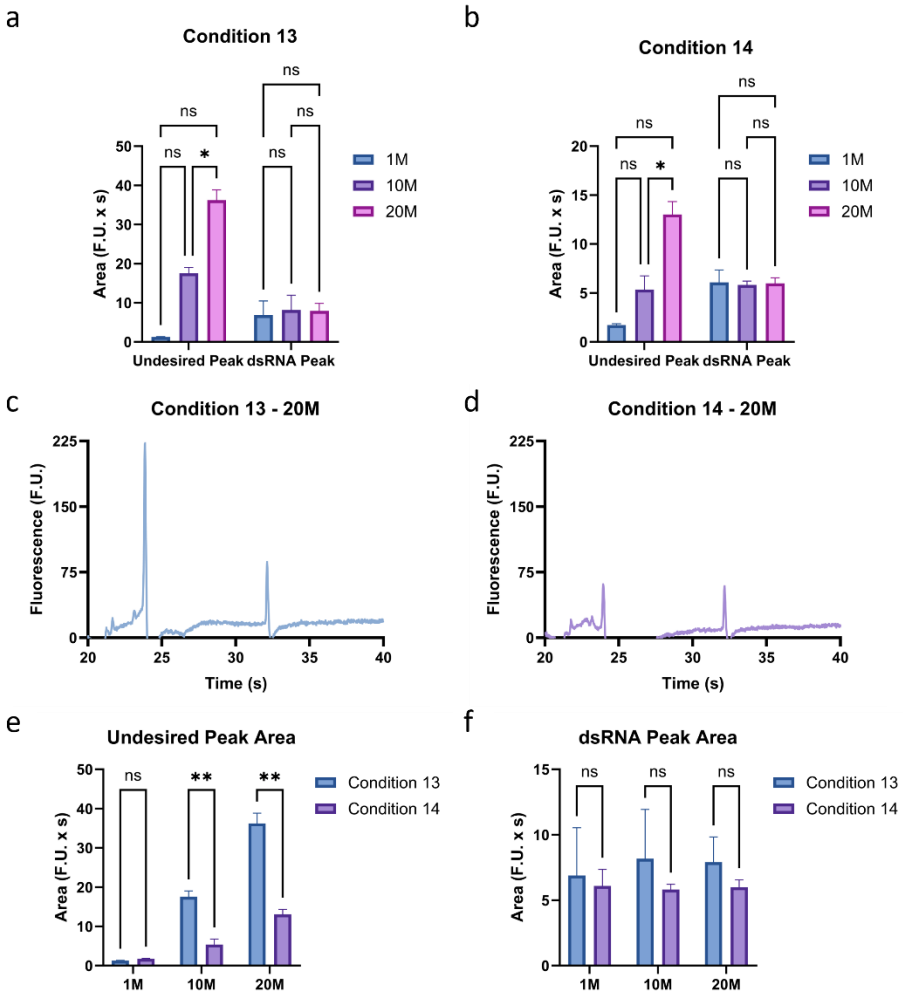

Figure 6.4: Comparison of conditions 13 and 14 at 1M (10 ng/µL), 10M (100 ng/µL), and 20M (200 ng/µL) of the 1,198 nt sample with the 700 bp dsRNA concentration kept constant at 5 ng/µL. Summarized areas of the undesired and dsRNA peaks using (a) condition 13 and (b) condition 14. Electropherograms of (c) condition 13 and (d) condition 14 at 20M. Summarized areas of the (e) undesired peak and the (f) dsRNA peak for conditions 13 and 14. "*" indicates the significance level in the difference (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$), while "ns" indicates no statistical difference. This figure was created using GraphPad Prism.

6.4.3 Validation of the Method

Once the treatment conditions were optimized (1/8 × buffer, 10 × enzyme), its performance was validated in terms of sensitivity (Figure S6.3) and length dependency (Figure 6.4). To assess the LOD, a 700 bp dsRNA sample was analyzed at 1.0, 2.5, and 5.0 ng/µL, and the peak area and height of each sample were used to determine the LOD using Equations 1 and 2 (Figure S6.3). A strong correlation can be observed between peak area and concentration (Figure S6.3a). No observable difference was observed between the area- and height-based estimations of the LOD, suggesting strong confidence in the LOD of the method estimated to be 0.32 ± 0.07 ng/µL of treated dsRNA. Next, to assess the length-dependency of the assay, it was tested with different dsRNA and mRNA lengths, specifically dsRNA of lengths 700 bp and 1,800 bp and mRNA of lengths 818 nt, 1,198 nt, 1,913 nt, 3,406 nt, and 4,451 nt (Figure 6.4). More specifically, Sample 1 contained 700bp dsRNA and 818 nt mRNA, Sample 2 contained 700 bp dsRNA and 1,198 nt mRNA, Sample 3 contained 700 bp dsRNA and 1,913 nt mRNA, Sample 4 contained 700 bp and 1,800 bp dsRNA and 1,913 nt mRNA, Sample 5 contained 700 bp and 1,80 0bp dsRNA and 3,406 nt mRNA, and Sample 6 contained 700 bp and 1,800 bp dsRNA and 4,451 nt mRNA. The concentration of dsRNA and mRNA in all samples was 1 ng/µL and 200 ng/µL, respectively. For all samples tested, there was a complete digestion of the mRNA peak, and, as can be highlighted in Figure 6.4, the dsRNA migration time was preserved for both the 700 and 1,800 bp fragments. Notably, this method can identify different dsRNA impurities in the same sample without interference.

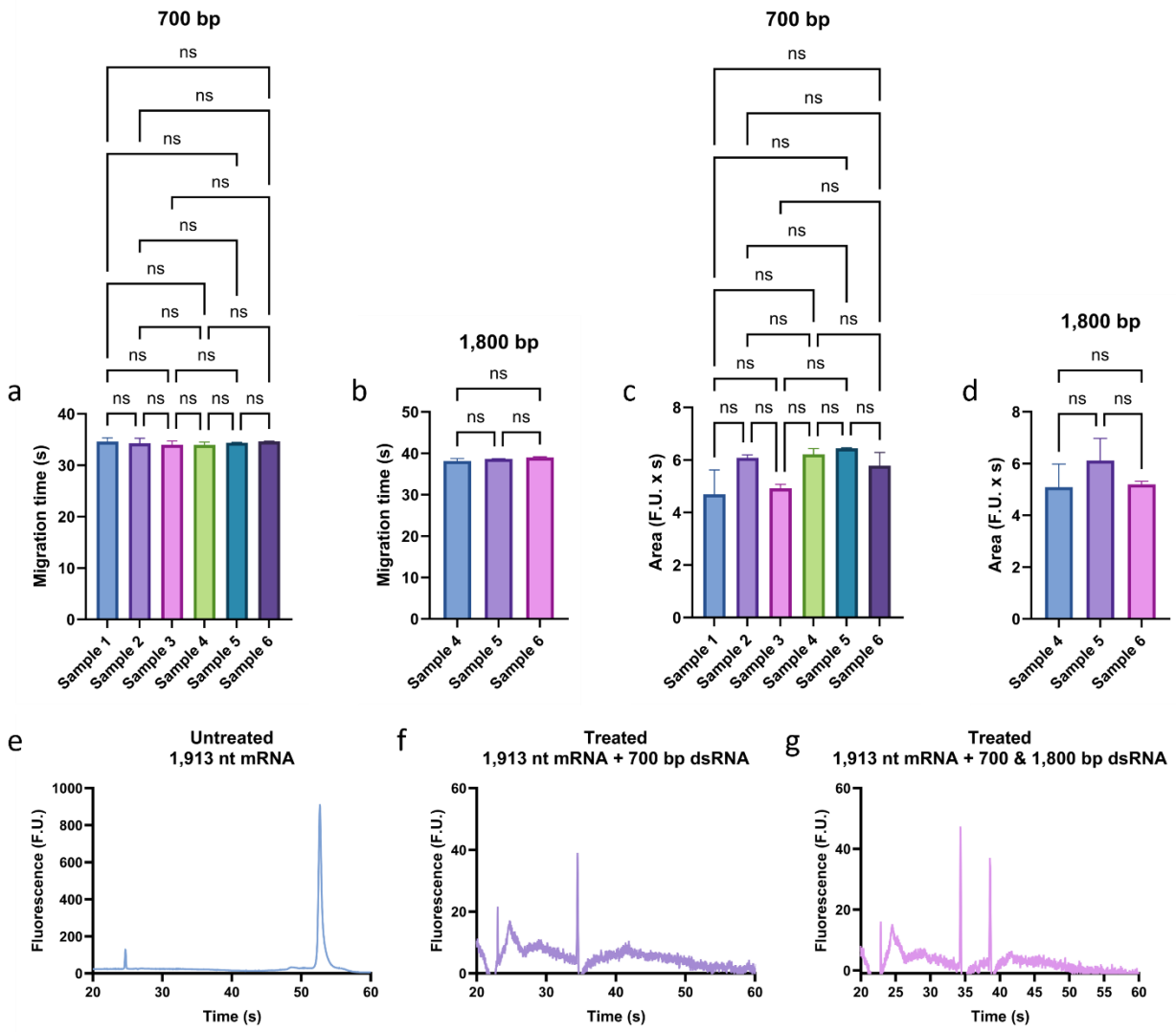

Figure 6.5: Assessing the robustness of the method against different mRNA lengths at a concentration of 200 ng/µL spiked with a 700 bp dsRNA or a 700 bp and a 1,800 bp dsRNA each at a concentration of 1 ng/µL. Since all mRNA peaks were dully digested by the enzyme, the dsRNA fragments were analyzed to determine whether the treatment affects the migration and integrity of the fragments that are not targeted by the enzyme. Migration time of (a) the 700 bp and (b) the 1,800 bp dsRNA for the samples containing each fragment. Peak area for (c) the 700 bp and (d) the 1,800 bp dsRNA for the samples containing each fragment. Representative electropherograms of (e) the untreated 1,913 nt mRNA, (f) the treated 1,91 3nt mRNA and 700bp dsRNA, and (g) the treated 1,913 nt mRNA, 700 bp dsRNA, and 1,800 bp dsRNA. Panels f and g highlight the complete disappearance through digestion of the 1,913 nt mRNA peak. "*" indicates the significance level in the difference (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$), while "ns" indicates no statistical difference. This figure was created using GraphPad Prism.

Once it was confirmed that the integrity of the dsRNA fragments was not affected, we tested the ability of the method to estimate the length of dsRNA impurities. To this end, once again a 1,913

nt fragment was spiked with two dsRNA fragments to simulate contaminants, however, this time the dsRNA fragments were added at a concentration of 0.25% relative to the mRNA concentration (Figure 6.5). Once the dsRNA fragments were identified, a ladder combined in-house was used to determine the length of the dsRNA impurities with an average error of 8%, ranging from 4-11%, with no statistical difference from the expected value. It must be noted that while the error is within our desired threshold, the ladder used to generate the curve that was used to estimate the lengths was composed of non-chemically modified dsRNA, while the samples were spiked with pseudouridine chemically modified dsRNA. This suggests that while their mobilities may differ slightly, a non-modified ladder can be used to estimate the length of pseudouridine modified dsRNA, and potentially other chemical modifications.

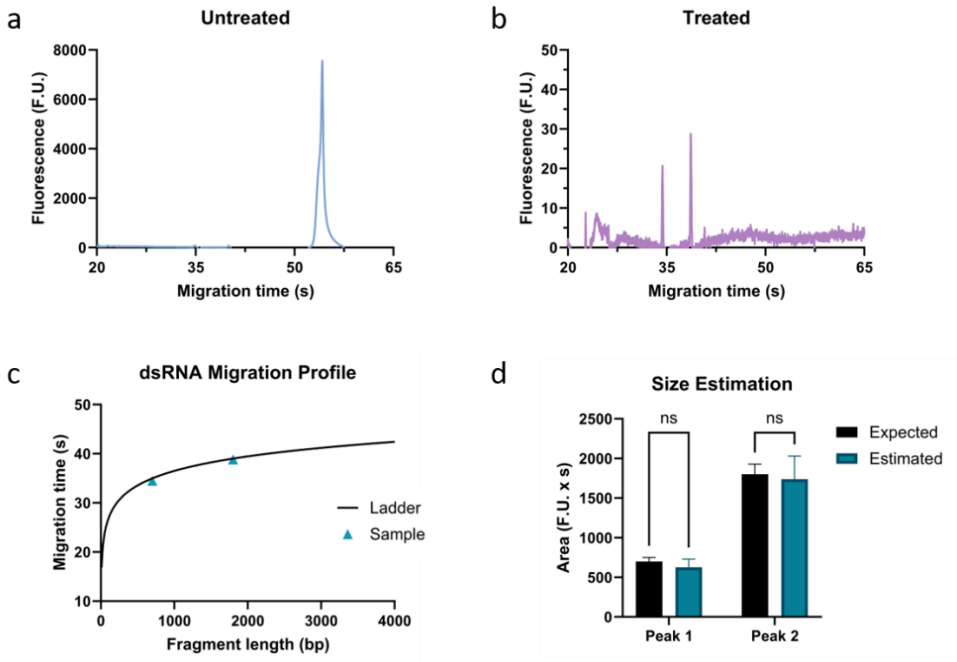

Figure 6.6: Assessment of the ability to estimate the dsRNA impurity fragment length from a sample containing a 700 bp and a 1,800 bp dsRNA fragment at a concentration of 0.5 ng/µL each and a 1,913 nt mRNA fragment at a concentration of 10 ng/µL (untreated) and 200 ng/µL (treated). The untreated electropherogram is for illustrative purposes with the same dsRNA concentration as the treated but an mRNA concentration 20 × lower due to the undigested mRNA concentration maximum loading capacity of the chip, which prevents us from loading more than 13 ng/µL[173] of untreated mRNA. Electropherograms of the (a) untreated and (b) treated samples. (c) Plot highlighting the logarithmic fit for the migration of the

ladder peaks (80, 150, 300, 500, 4,001 bp), $y = 9.82\log(x) + 7.04$ and $R^2 = 0.99$. (d) Comparison of the expected (black) and estimated (teal) fragment length of the two dsRNA peaks using the dsRNA ladder fit, showing no statistical difference between the expected and estimated values for both peaks. Figure created using GraphPad Prism.

6.5 Conclusion

This study presents a novel method for characterizing dsRNA impurities in mRNA formulations. Microfluidic electrophoresis offers a rapid, high-throughput analytical tool for the fragment-based analysis of both dsRNA and mRNA fragments, enabling the determination of their lengths through ladders. However, as discussed in our previous study,[173] long dsRNA and mRNA molecules can differ significantly in electrophoretic mobility, requiring nucleic acid-specific ladders for size determination. Therefore, since, in addition to the residual dsRNA, truncated or degraded mRNA fragments may also be present in the sample,[221] it is crucial to determine the identity of the dsRNA peaks before assessing their size.

To this end, we developed an enzymatic identification method compatible with our microfluidic system that uses S1 nuclease to digest all single-stranded fragments in the sample, only leaving behind the dsRNA fragments. While the standard protocol for the enzyme treatment yielded acceptable results, we optimized the concentrations of buffer and enzyme used in the protocol to increase the sensitivity and robustness of accepting a wide range of mRNA concentrations. Upon optimizing these conditions, the LOD of the method for dsRNA post-treatment was calculated to be 0.32 ± 0.07 ng/μL, 8.8-fold greater than using the SOP, and the maximum loading concentration of mRNA tested in this study was 200 ng/μL. Therefore, the lowest percentage of dsRNA contaminant experimentally tested in an mRNA sample here was 0.25%; however, based on the LOD, we expect the system to be capable of detecting percentages as low as 0.16% when 200 ng/μL of mRNA are loaded. It should be noted that it is also possible that higher concentrations of

mRNA may be analyzed using the proposed method, further lowering the percentage of impurities that can be detected, but it was not directly assessed during this study.

The focus of this paper was primarily the development of an enzymatic identification method for residual dsRNA in mRNA samples. Once the peak was identified, we were able to use the line of best fit yielded by a ladder we combined in-house to estimate the length of the impurities within 8% of the expected length. While in this study, a ladder had to be run on the same plate as the sample and the migration over length of the ladder peaks had to be fit to generate the curve used to estimate the sizes, we are currently working on developing a more streamlined sizing method. Using experimental data, we are developing a physics-informed neural network model capable of predicting dsRNA and mRNA electrophoretic mobility under different gel concentrations using our microfluidic platform.

Requiring only 10-15 µL of sample, the proposed method has an analytical run time of 1 min/sample post enzyme treatment. Since the analytical platform is compatible with 96- and 384-well plates, and the enzyme treatment can be conducted directly on the well plate, this method is considered high throughput, requiring human intervention during the microfluidic analysis. The low sample requirements and high throughput nature of the method make it highly suitable for the batch-to-batch monitoring of residual dsRNA and for the mRNA process development optimization in the biopharmaceutical industry.

6.6 Author Contributions

Conceptualization, A.C.D.P.; data curation, A.C.D.P., M.V., and N.L.; formal analysis, A.C.D.P., and M.V.; funding acquisition, A.T.; investigation, A.C.D.P., N.L., S.S., and M.B.F.; methodology, A.C.D.P., and M.V.; project administration, A.C.D.P., and A.T.; resources, A.T.; supervision, A.C.D.P., and A.T.; validation, A.C.D.P., M.V., N.L., S.S., and M.B.F.; visualization, A.C.D.P.,

M.V., and N.L.; writing – original draft, A.C.D.P., and M.V.; writing – reviewing and editing, A.C.D.P., M.V., N.L., S.S., M.B.F., and A.T.

6.7 Acknowledgements

This work was in part supported by Revvity's research grant to Brown University. A.T. is a paid scientific advisor/consultant and lecturer for Revvity.

6.8 Conflict of Interest

The authors declare that there are no conflicts of interest.

6.9 Supporting Information

Table S6.1: Conditions tested during the factorial analysis

Table S6.1: Factorial design of experiment with three parameters - mRNA, buffer, and enzyme concentrations - at three levels. The mRNA concentrations were 1/2, 1, and 2 × the initial concentration of 10 ng/µL of a 1,198 nt sample, the buffer concentrations were 1/16, 1/8, and 1/4 × the initial concentration of 9% v/v of the stock, and the enzyme concentrations were 2, 10 and 50 × the initial concentration of 0.30% v/v of the stock. For all conditions, the 700 bp dsRNA sample concentration was kept constant at 5 ng/µL.

Condition	dsRNA Concentration (ng/µL)	mRNA Concentration (× 10 ng/µL)	Buffer Concentration (× 9% v/v)	Enzyme Concentration (× 0.30% v/v)
Untreated	5	1	0	0
1	5	1/2	1/16	2
2	5	1/2	1/8	2
3	5	1/2	1/4	2
4	5	1/2	1/16	10
5	5	1/2	1/8	10
6	5	1/2	1/4	10
7	5	1/2	1/16	50
8	5	1/2	1/8	50
9	5	1/2	1/4	50
10	5	1	1/16	2
11	5	1	1/8	2
12	5	1	1/4	2
13	5	1	1/16	10
14	5	1	1/8	10
15	5	1	1/4	10
16	5	1	1/16	50
17	5	1	1/8	50
18	5	1	1/4	50
19	5	2	1/16	2
20	5	2	1/8	2
21	5	2	1/4	2
22	5	2	1/16	10
23	5	2	1/8	10
24	5	2	1/4	10
25	5	2	1/16	50
26	5	2	1/8	50
27	5	2	1/4	50

Figure S6.1: Comparison of the dsRNA and undesired peak areas yielded by the factorial analysis

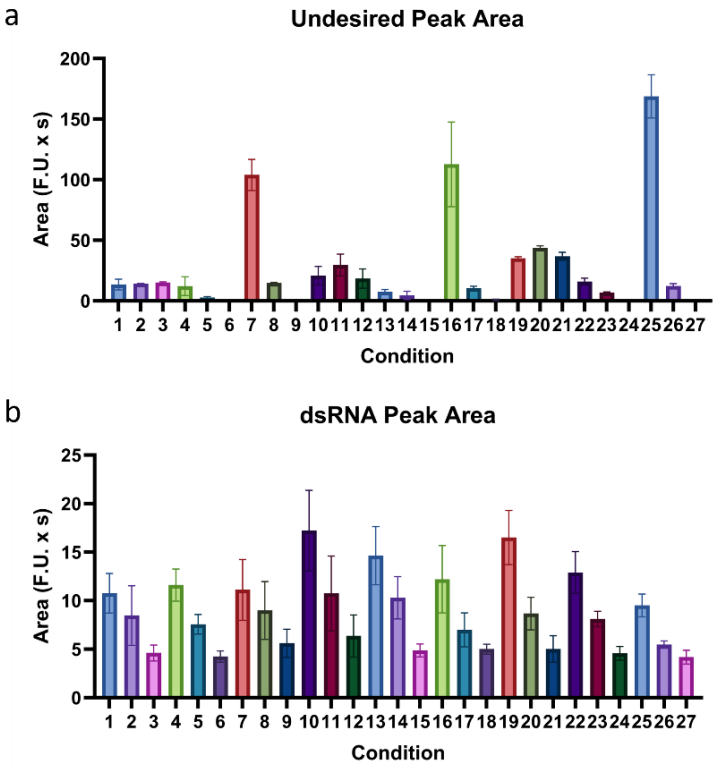

Figure S6.1: Comparison of the (a) undesired peak area and (b) dsRNA peak area yielded by the 27 conditions tested. From here, the best 10 conditions based on lowest undesired peak areas were further explored in the main text.

Figure S6.2: Factorial analysis of the buffer, enzyme, and mRNA concentration experiment

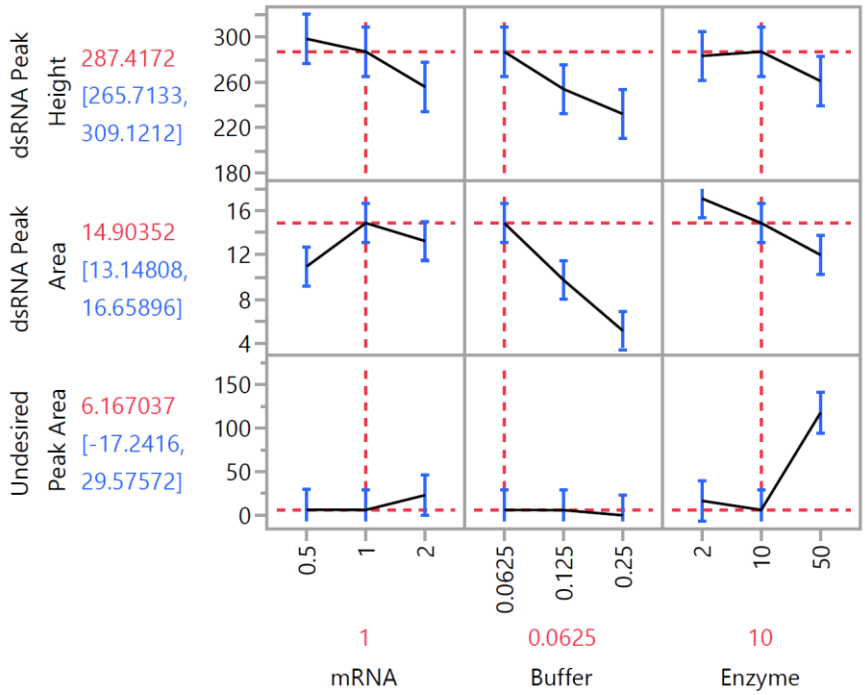

Figure S6.2: Factorial analysis changing three parameters - mRNA, buffer, and enzyme concentrations - at three levels. During the design of experiments, the mRNA concentrations were 1/2, 1, and 2 × the initial concentration of 10 ng/µL, the buffer concentrations were 1/16, 1/8, and 1/4 × the initial concentration of 9% v/v of the stock, and the enzyme concentrations were 2, 10 and 50 × the initial concentration of 0.30% v/v of the stock. For the statistical analysis, a factorial analysis was performed on JMP Pro 17, with the goal of maximizing the dsRNA Peak Height and the dsRNA Peak Area and minimizing the Undesired Peak Area. The figure was generated using JMP Pro 17.

Figure S6.3: Assessment of the dsRNA limit of detection

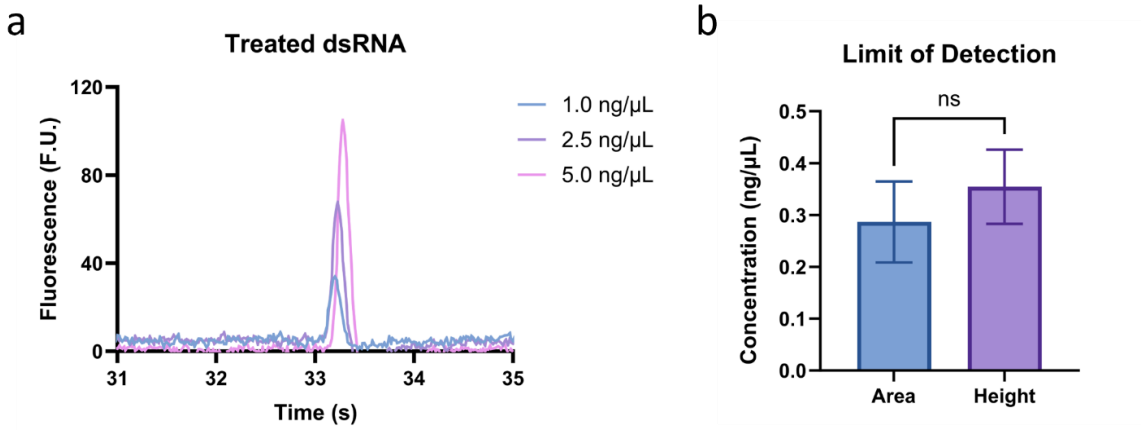

Figure S6.3: Assessment of the LOD of the system post enzymatic treatment, estimated by analyzing a 700 bp dsRNA sample at 1.0, 2.5, and 5 ng/µL. (a) Electropherogram of the three dsRNA concentrations. (b) Comparison of the average LOD estimated from the three samples using Equations 1 (area) and 2 (height). "*" indicates the significance level in the difference (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$), while "ns" indicates no statistical difference. This figure was created using GraphPad Prism.

Chapter 7: Empirical Analysis and Physics-Informed Neural Network-Based Prediction of mRNA and dsRNA Electrophoretic Behavior

---

Chapter 7 has been submitted as an original research article.

Coll De Peña, A.,[1] Li, N.S.,[1] Vaduva, M.,[2] Goswami, S.,[3] Tripathi, A.,[1] *Empirical Analysis and Physics-Informed Neural Network-Based Prediction of mRNA and dsRNA Electrophoretic Behavior*, **in review**.

[1]Center for Biomedical Engineering, School of Engineering, Brown University, Providence, RI, USA

[2]Department of Molecular Biology, Cell Biology, and Biochemistry, Division of Biology and Medicine, Brown University, Providence, RI, USA

[3]Applied Mathematics, Brown University, Providence, RI, USA

7.1 Abstract

RNA-based therapeutics are currently at the forefront of the biopharmaceutical industry because of their safety, efficacy, and shortened time from disease discovery to therapy development. Microfluidic electrophoresis provides a great analytical platform to analyze nucleic acids in unprecedented detail. However, while DNA has been studied extensively within microfluidic systems, there is limited data available for RNA, particularly of chemically modified molecules, such as those used in the COVID-19 mRNA vaccines, and for long double-stranded RNA molecules, which may accompany, intentionally or as a by-product, RNA therapeutics. To this end, this study focused on the empirical microfluidic electrophoretic analysis of double- and single-stranded RNA, non-modified and pseudouridine-modified, at varying gel concentrations. It then compared the findings to the electrophoretic mobility models in the literature. This work was then complemented with data-driven and physics-informed neural networks that successfully predicted the migration time and length of different RNA molecules with an average error of 15.8% for the data-driven model and 0.77% for the physics-informed model. The low error in the physics-informed neural networks opens the doors to the electrophoretic characterization of molecules, even beyond RNA, without the need for extensive experimental data.

7.2 Introduction

The advent of microfluidic electrophoresis as a nucleic acid analytical tool has allowed scientists to study the migration of DNA and RNA at an unprecedented level of detail. Despite the promise this platform brings due to its characteristic short runtime, high resolution, and increased sample throughput,[6,51] the migration patterns of these molecules in microfluidic conditions have not been extensively studied. Furthermore, ideal experimental conditions for nucleic acids of various lengths have not been clearly defined. Studying the biophysical interactions between migratory nucleic acids and their medium is vital for developing nucleic acid purity and integrity analysis assays, such as mRNA vaccine quality assessment.[170,173] It is also crucial for developing more complex assays to resolve similar molecules. Such advancements may improve the characterization of exosome genomes[238–241] and quality control assays for multivalent nucleic acid-based vaccines.[242–244] Understanding the mechanisms behind nucleic acid migration is also crucial for troubleshooting assays, allowing researchers to identify experimental conditions that better suit their parameters and desired outcomes.

The current models describing nucleic acid electrophoretic mobility are differentiated according to the relative sizing between the pore size of the semi-dilute polymeric network and the nucleic acid size, as defined by the radius of gyration (R_g).[245] The main models include the Ogston model, describing DNA molecules with R_g smaller than the pore size, and the Biased Reptation with Fluctuation (BRF), describing DNA molecules with R_g greater than the pore size. The BRF model is further differentiated into two scenarios: reptation without orientation and reptation with orientation. The Ogston model assumes the DNA is a spherical object moving through a sieve driven by the electric field.[245,246] In this model, mobility is proportional to the exponential of the negative concentration of the polymer solution.[245,246] According to the BRF, mobility scales as 1/N

for short chains and levels off for large sizes and/or high electric fields.[245,247] Each model has respective limitations and alternative modifications have also been made to better describe nucleic acid movement in capillary electrophoresis.[248–251] While there has been abundant research in the electrophoretic separation of DNA, much less work has focused on the separation and mobility models of RNA, especially long nucleoside-modified mRNA and double-stranded RNA (dsRNA).[252–254]

This study aims to explain the electrophoretic mobility of different RNA molecules in microfluidic systems and understand the underlying physical principles that govern the migratory patterns of differently sized RNA in varying concentrations of semi-dilute polymer solutions. Given their clinical relevance in mRNA vaccines and therapies, a focus is placed on the mobility of both single- and double-stranded RNA and the potential impact of nucleoside modifications.[2,235,236] To the authors' knowledge, this is the first time the electrokinetic of dsRNA fragments, especially that of longer length and nucleoside-modified RNA, has been studied in microfluidic capillary electrophoresis. With the rise in RNA research, the mobility of RNA with pseudouridine modifications, which enhance RNA stability and decrease their immunogenic response,[236,237] will be important to characterize. Additionally, immunogenic dsRNA, intentional or residual, will be a critical component in future vaccine and alternative therapy research and development. This paper also aims to provide predictive modeling of the electrophoretic mobility of single- and double-stranded RNA of varying lengths under different conditions using artificial neural networks (ANNs) to provide guidelines for future assay development, diagnostic protocols, quality control platforms, and genetic material differentiation.

ANNs are a class of machine learning tools[255–257] inspired by the structure and behavior of biological neural systems based on research on artificial neurons.[258] Since its introduction, it has

proven to be a powerful and versatile tool for problems where the solutions are not clearly known or insufficient information has been given for the relationship between the inputs and the output. It can determine complex linear and non-linear relationships. ANN recognizes the patterns in a series of input and output values, and using the acquired 'knowledge', it then predicts the unknown output values for a given set of input values.[256] In this study, our goal is to employ ANNs to distinctly solve the following two tasks: task 1 - to predict the migration time of single- and double-stranded RNA molecules given the experimental conditions and physical properties of the molecules, and task 2 - to predict the length of single- or double-stranded RNA molecules given the experimental conditions and migration time of the molecule. To develop efficient frameworks for completing the tasks discussed above, we investigate two specific paradigms of ANNs:

Framework 1: Data-driven:[259,260] In this setup, the ANNs are trained on labeled datasets. The model is trained to detect underlying patterns and relationships between the input and the output data, enabling it to yield accurate predicted results when presented with unseen input data. Data-driven frameworks are dependent on the volume and the quality of the labeled dataset. Often, it fails to ensure that the physics of the problem is satisfied.

Framework 2: Physics-driven:[261,262] In this setup, we train the ANNs with the governing physical equation alongside some labeled experimental/computational data. This method is highly efficient in scenarios where we may have access to a limited number of training samples (i.e., we work in a small data regime).

To tackle the tasks, we converted them into a regression problem and used each framework to solve it. Using the data-driven network, framework 1, we obtained an average error of 12.68% for task 1 and 18.92% for task 2. In contrast, using the physics-driven network, framework 2, which is more suitable for smaller data sets, we obtained an average error of 0.44% for task 1 and 1.1%

for task 2, a significant improvement from framework 1. These results highlight how ANNs, in particular physics-informed neural networks (PINNs), can be used to increase the understanding of physical behavior in biological samples, such as their electrophoretic mobility, and to support the development of analytical methods by significantly decreasing the number of experiments needed.

7.3 Materials and Methods

7.3.1 Samples and Sample Preparation

The ssRNA (catalog #N0364S) and dsRNA ladders (catalog #N0363S) used in this study were purchased from New England Biolabs (New England Biolabs, Ipswich, MA). The ssRNA ladder contains fragments of 50 nt, 80 nt, 150 nt, 300 nt, 500 nt, and 1,000 nt, while the dsRNA ladder contains fragments of 21 bp, 30 bp, 50 bp, 80 bp, 150 bp, 300 bp and 500 bp. The non-chemically modified, 4,001 bases long mRNA and dsRNA samples were custom ordered from Genewiz (Genewiz Genomics Headquarters, South Plainfield, NJ). The pseudouridine chemically modified mRNA (818, 1,198, 1,913, 3,406, and 4,451 nt) and dsRNA (700 and 1800 bp) fragments, and the 700 and 1,800 bp non-chemically modified dsRNA fragments were purchased from CATUG Biotechnology (CatPure™; Cambridge, USA). Prior to use, the ssRNA and dsRNA ladders were diluted 1:5 in 1X TE buffer, and the custom mRNA and dsRNA fragments were diluted to 5 ng/μL in 1X TE buffer.

7.3.2 Microfluidic Measurements and Analysis

For the microfluidic electrophoretic characterization of the samples, the LabChip GXII Touch platform (Revvity, Waltham, MA) was used to control the electric fields used to control the migration of the samples for analysis. The platform was combined with a custom RNA microfluidic chip, SYTO 61, poly(N, N-dimethyl acrylamide) (PDMA) solution, and lower

marker, all provided by Revvity, as described in our previous study.[173] To analyze the electrophoretic behavior of the samples at different gel concentrations, the stock PDMA solution was diluted with a gel diluent (Revvity) that maintained the conductivity constant while the gel concentration changed. After diluting the gel to the desired concentration, it was mixed with 2.34% v/v of the fluorescent stain and spun down according to the instructions from the provider. The chip was loaded with the gel-dye mixture and the lower marker for electropherogram alignment during analysis.

Once the samples were diluted to the desired concentration, 10-15 μL were loaded onto a 384-well plate, and the well plate and chip were loaded onto the platform. Independent of gel concentration, a script containing the same loading, injecting, and separation voltages, which have been described in our previous study, was used for all experiments.[173] However, the separation time was increased depending on the gel concentration to ensure all peaks were captured. After the script was run, the LabChip Reviewer software (Revvity) was used to visualize the electropherograms.

Unless otherwise specified, all experiments were conducted in 2-3 experimental repeats with 2-3 instrumental repeats, yielding 6-9 data points per condition and sample tested. The statistical analyses were conducted using GraphPad Prism 9.4.1 (681), and the significance was by a Tukey post hoc test with a confidence interval of 95%; *$p<0.05$, **$p<0.01$, ***$p<0.001$, ****$p<0.0001$. GraphPad Prism 9.4.1 was also used to generate the non-linear regression fits reported across the study.

7.3.3 Theoretical Operating Regime

Unlike traditional slab gel electrophoresis, where the cross-linked polymers create fixed pores through which analytes can migrate, capillary electrophoresis in entangled polymers relies on

transient pores created by entanglement points that undergo a "random walk". The pore size of various gel concentrations can be approximated by the blob size (ξ_b) and is calculated by:[245]

$$\xi_b = 1.43R_g(\frac{c}{c^*})^{-(1+a)/3a}, \quad (7.1)$$

where c is the polymer concentration, c^* is the entanglement threshold for the polymer, a is an exponent of the Mark-Houwink equation, and R_g is the radius of gyration of the polymer calculated by:

$$R_g \approx \sqrt[3]{\frac{[\eta]M_r}{6.2N_A}}, \quad (7.2)$$

where [η] is the intrinsic viscosity of the polymer solution at the concentration, M_r is the polymer molecular mass, and N_A is Avogadro's number. Using the Mark-Houwink coefficients established previously for PDMA[211] and intrinsic viscosity found through linear interpolation of PDMA solutions of various molecular weights,[263] an operating effective pore size of 37 nm, 21 nm, 15 nm, 12 nm, and 10 nm were established for 1%, 2%, 3%, 4%, and 5%, respectively.

When analyzing mobility, it is important to examine the relationship between the pore size of the sieving matrix and the radius of gyration of the nucleic acid. Defined as the average distance squared between different parts of the object and its center of mass, this measurement provides information regarding the average shape of the nucleic acid that end-to-end distance does not, which is essential to characterize as electric fields can induce different molecular conformations.[264] The radius of gyration of nucleic acids can be approximated by:[265]

$$R_g^2 = \frac{1}{3}pL[1 - \frac{p}{L} + \frac{p}{L}e^{-L/p}] \quad (7.3)$$

where p is the persistence length, a measurement of polymer chain stiffness, and L is the length of the polymer fragment. Values of 64 nm and 2 nm for the persistence length of dsRNA and ssRNA were used for calculations based on respective averages of current literature reporting.[266–268] Approximating the transition from Ogston-like motion to BRF as the size where the R_g of the nucleic acid is equal to the effective pore size, we predicted that all dsRNA samples would operate under reptation with the possible exception of 80 bp under 1% gel. For ssRNA, this transition point is expected to occur from approximately 54 – 2150 nucleotides for 1 – 5% gel. However, we acknowledge the true transitions may vary due to the heavy reliance of the formula on persistence length, which has been reported with large ranges, and the approximations made in formulaic calculations of pore size.

7.4 Results and Discussion

7.4.1 Electrokinetic Behavior of dsRNA and ssRNA

The electrokinetic behavior of migratory nucleic acids is mainly affected by the relative sizing of the molecule and the matrix pore size, the strength of the electric field, and the interactions between the matrix and nucleic acid during electrophoresis. This was demonstrated by the raw migration trends of dsRNA and ssRNA, which were reflective of typical manipulation of gel concentration and fragment size. Fragments of the same size migrated slower as gel concentration increased due to increased viscosity and decreased pore size; however, resolution between longer fragments of nucleic acids significantly decreased at lower gel concentrations (Figure S7.1). Additionally, the migration time of dsRNA and ssRNA of small fragment lengths (< 300 bases) were relatively similar and within a 10% difference of each other for respective gel concentrations. As fragment size increased, the faster migration of dsRNA relative to ssRNA became more prominent, especially at higher gel percentages, where there was approximately a 3%, 12%, 51%, 53%, and

71% difference in migration time between the 4,001 base fragment size of the two nucleic acids in 1 to 5% gel, respectively. It is also interesting to note that in the observed size ranges, as gel concentration increased from 1% to 5%, the migration difference between the largest and smallest fragment (150 and 4,001 fragment length) of dsRNA was less pronounced, with a 28% to 43% difference, than for ssRNA, which had a 28% to 100% difference. This can also be seen visually through how the overall shape of the trend remains constant between the gel percentages as fragment size increases for dsRNA, but the data points span outwards as fragment size increases (Figure S7.1a). In a previous comparison of RNA, ssDNA, and dsDNA, it was also found that single-stranded nucleic acids had better separation and higher resolution over a larger range of sizes compared to dsDNA.[269] This may be due to the greater flexibility of single-stranded molecules and a tendency to form secondary and/or tertiary structures that could cause interactions with the sieving matrix.

The electrophoretic mobility of RNA of varying sizes and type were calculated and demonstrated in Figure 7.1. The overall sigmoidal curve demonstrated by the double logarithmic mobility vs. size plots agrees with previous DNA and RNA capillary electrophoresis separation studies.[211,270] However, this shape is much less evident for ssRNA (Figure 7.2b). Transitions were demarked upon visual inspection with solid gray lines between Ogston-like sieving (Regime I), Reptation without orientation (Regime III), and Reptation with orientation (Regime IV) with patterns outlined by previous studies.[271,272] Regime II, which was only prominent in dsRNA, marks the regime described by Heller,[211] where R_g may be greater than pore size, but the dsRNA is still too stiff to reptate.

According to the Ogston model,

$$\frac{\mu}{\mu_0} = e^{-K_R \times C} \quad (7.4)$$

where μ_0 is free solution mobility, K_R represents the Retardation coefficient, and c is gel concentration, linear Ferguson plots (logarithm of mobility as a function of gel concentration) would indicate RNA separation within the Ogston sieving regime. RNA is presumed to migrate through the gel matrix in a globular confirmation in a random array of cylindrical obstacles in this region. Interestingly, for dsRNA, a straight linear line through the data points could be argued for all studied fragment sizes (Figure 7.1c). Comparing the straight line fit for 80 bp ($Y = 2.025\ e^{-0.1024c}$; $R^2 = 0.96$) and 4001 bp ($Y = 1.523\ e^{-0.1524}$; $R^2 = 0.96$), it appears the Ogston model does not fit the smaller strands any better than the longer fragment, where mobility seems to fail. As expected, the retardation coefficient increases with increasing fragment lengths while the free solution mobility decreases.

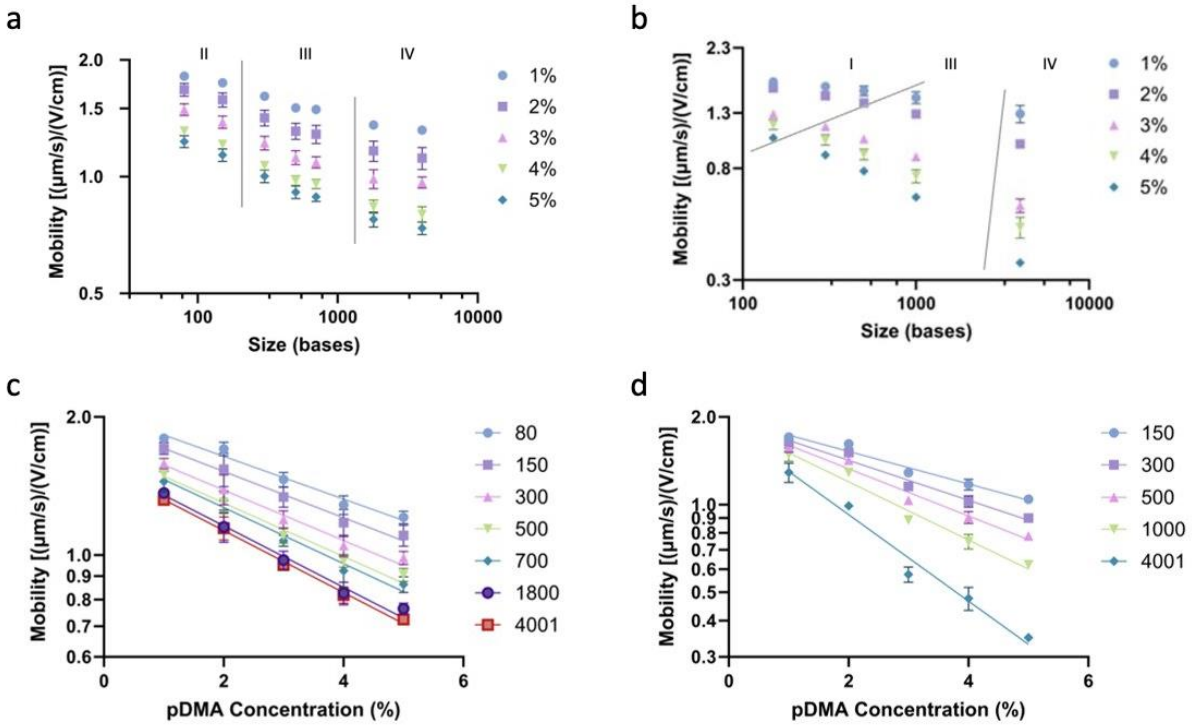

Figure 7.1: Double-logarithmic mobility vs. size plots for (a) dsRNA and (b) ssRNA at polymer concentrations of 1%, 2%, 3%, 4%, and 5%, at an electric field strength of 417.4 V/cm where the lines are used to semi-qualitative define the different regimes. The same data is visualized as Ferguson plots for (c) dsRNA and (d) ssRNA and are exponentially fitted to demonstrate potential reptation motion.

While it is surprising that the model has a similar fit for the whole range of dsRNA, it does confirm that fragment lengths as small as 80 bp may be too large to sieve through the pores without stretching. For ssRNA, the data points noticeably deviated from a straight line fit of negative slope, with the possible exception of ssRNA of 150 bases (Figure 2d). These inconsistencies demonstrate that the Ogston mobility model may not be an accurate method of describing the mobility of RNA, especially for single-stranded RNA. It may also indicate that more of the Regime I indicated in Figure 7.2b may actually fall under reptation motion. This finding is similar to previous studies that have also found the Ogston-like sieving mechanism to not accurately describe the mobility of ssRNA but instead be better for dsDNA and ssDNA.[269] Our study demonstrates that this may also apply to dsRNA, meaning that it is not the helical structure of short nucleic acids but the

composition of the strands or other formed structures that contribute more to conditions described by Ogston-like motion. This may be due to the many conformations RNA can take on, which may not necessarily be spherical and instead have loops or other secondary/tertiary structures, and with a persistence length much less than that of dsRNA, the effect of the electric field on the ssRNA and resulting stretching may be even more pronounced.

Three distinct regions can be seen for dsRNA in Figure 7.1a, and with our pore size analysis and the poor fitting of the Ogston model, the first region of the data points may be better described as what Heller concludes as a transition region where R_g is greater than pore size, but the molecules are too still to reptate.[211] As gel concentration increases, this distinction is less prominent, and there seems to be a linear decrease in mobility in Figure 1a until a plateau in Region IV is reached. This contradicts what Heller noted, where Regime II was most expressed at higher polymer concentrations, and a smooth transition from Regime I to II was present at low concentrations. This may indicate that smaller fragment sizes of dsRNA favor stretching and uncoiling at higher gel concentrations due to smaller pore sizes and increased polymer density.

While the Ogston models don't seem to fit the shorter strands of the RNA very well, the Ferguson graphs (Figure 7.1c,d) do provide some insights into the relative free solution mobility of dsRNA and ssRNA in our system, being approximately 1.957 and 1.971 respectively, as approximated by the y-intercept of the equations fitted by the 150-base fragment. This demonstrates that ssRNA may have a slightly higher, but not necessarily significant, free solution mobility than dsRNA. This is surprising given the greater charge density but can be explained by the greater flexibility and potential compactness of ssRNA.

Unlike Ogston-like motion, the reptation model was developed for larger DNA due to the assumption that the spherical coil would be too large to fit through the pores of the matrix

undeformed and would instead migrate head-first in a "snake-like" motion through "tubes" formed by the polymeric pore networks.[273] Later improved by researchers such as Slater and Viovy, the model, while still challenging to utilize mathematically, was modified to account for larger sizes and high fields where separation fields as well as the dynamic nature of uncrosslinked polymer networks[274] and is represented by:[275]

$$\frac{\mu}{\mu_0} = \left\{ \left[\frac{\left(\frac{b}{l}\right)^2}{3N_k} \right]^2 + \left[\frac{2\varepsilon_k \left(\frac{b}{l}\right)^2}{5+2\alpha\varepsilon_k \left(\frac{b}{l}\right)^2} \right]^2 \right\}^{\frac{1}{2}} \tag{7.5}$$

where b is the pore size, l is the Kuhn length, N_k is the number of Kuhn segments, α is the limiting ratio of free mobility to in-gel mobility for large chains and large fields, and ε_k is the "molecular" reduced field.

Regime III of Figure 7.1a,b can be correlated with regions of reptation without orientation, as seen by the linear decrease on the double logarithm of mobility as a function of fragment length (Figure 7.1a,b) and validated by the linear region of Figure 7.2.

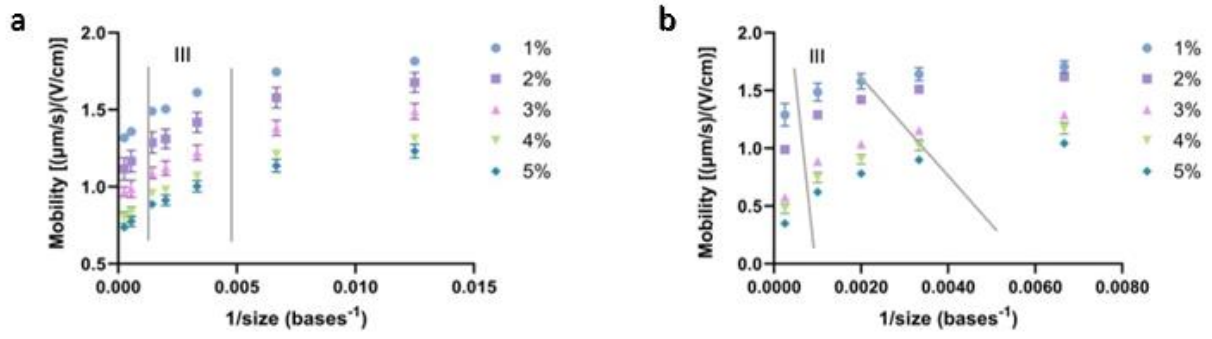

Figure 7.2: Mobility vs 1/size for (a) dsRNA and (b) ssRNA where the lines are used to semi-qualitative define the different regimes. The linear region demarked as Regime III represents reptation without orientation.

As Equation 5.26 from Viovy demonstrates, the first term dominates for molecules below a critical size ($N^* = N_k$), and mobility is inversely proportional to fragment size. However, as the N_k

becomes larger than N*, a plateau mobility is reached. The reptation with orientation Regime that describes separation failure is thought to be partly due to the electric forces leading long fragments to choose a tube of consecutive pores that do not follow a random walk-in space, and therefore resulting mobility is independent of size.[265] The reptation Regime has also been explained by other authors as countering the effects of increased charged residues against increased solvent friction due to large nucleic acid size, the latter of which can also be attributed to collisions and subsequent transient dragging of polymer chains.[248,249] This plateau can be seen around 1000 bp for dsRNA (Figure 7.1a) and seems to be reached a little later for ssRNA/mRNA in the same condition as shown in the region demarked Regime IV in Figure 7.1b. The higher critical sizes of single-stranded nucleic acids are in agreement with previous findings.[269] Similar to what was previously noted for ssDNA and dsDNA, this critical size seems to increase with decreasing polymer concentration for ssRNA but remained constant for dsRNA,[211,270] but this is hard to confirm due to gaps in data from sample limitations. However, other studies have found the transition between regimes to be dependent on solution concentration only for RNA and not for ssDNA, which is a potential explanation for the short-lived secondary structures that ssRNA can make, resulting in increased stiffness.[269] As expected, the resolving power for larger fragment sizes is poor in very low gel concentrations, but separation also fails earlier as gel concentration is increased, shown mainly by ssRNA/mRNA. However, there was a consistent increase in peak width as gel concentration increased for both dsRNA and ssRNA.

7.4.2 Comparison and Differentiation of dsRNA and ssRNA

Given that new developments in RNA therapeutics involve a combination of dsRNA and ssRNA/mRNA, it is essential to analyze potential differences in mobility patterns for identification. Given that the unique mix results in a preference for not denaturing the ssRNA due

to the possible unraveling of the double-stranded helix of dsRNA, our findings will provide insights into mobility that may be complicated by secondary structures that are typically linearized and subsequently reported in the literature.

Figure 7.1a,b demonstrates that the difference in mobility for a single fragment between different gel % is relatively constant for dsRNA but not for single-stranded RNA, which is consistent with the finding of Heller in terms of dsDNA and ssDNA.[211] To analyze the influence of pore size on the mobility of fragments of different lengths, a semilogarithmic plot (Figure 7.3) was graphed to directly assess the dependence of mobility on pore size for fragment sizes 500 and 4001, which should respectively fall under reptation without orientation and reptation with orientation. The dependence of dsRNA mobility on pore size was similar for both fragment sizes, with slopes of approximately 0.47. However, logarithmic trends did not seem to fit our ssRNA data, although the R^2 is similar to the trends for dsRNA. There was a considerable difference between the slope of the 500 nt trend (0.64) and that of 4001 nt (0.76).

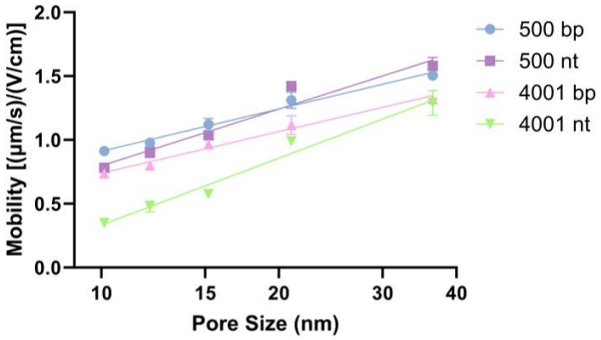

Figure 7.3: Dependence of RNA electrophoretic mobility on pore size.

To analyze the potential difference in mobility between dsRNA and ssRNA in terms of gel concentration, the mobility of both RNA types was plotted against fragment size for each gel condition (Figure 7.4). Figure 7.4a demonstrates that at low gel concentrations (1% and 2%), the mobility of ssRNA and dsRNA are very similar across all fragment sizes (Figure 7.4a), but as gel

percentages increase, there is a clear increase in dsRNA mobility compared to that of ssRNA (Figure 7.4b). This difference becomes more pronounced as fragment size increases. The mobility of dsRNA and ssRNA in 1% gel is approximately equal (<5% difference). For gel concentrations greater than 2%, the difference in mobility increases between dsRNA and ssRNA as fragment size increases above ~ 500 bases for each gel percentage (Figure 7.4b), before which the mobility difference is less than 10%. Additionally, as gel percentages increase from 1% to 5%, the difference in mobility for the largest recorded fragment also increases from 2.2% to 60.5%. This trend was also seen when analyzing raw migration time, demonstrating the potential manipulation of higher gel percentages to increase the separation between dsRNA and ssRNA.

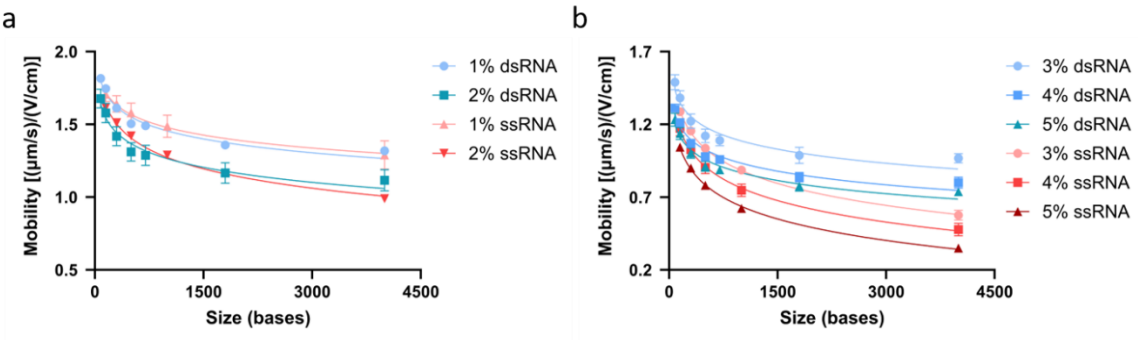

Figure 7.4: Mobility vs fragment size of dsRNA and ssRNA in (a) 1-2% gel and (b) 3-5% gel.

7.4.3 Impact of Chemical Modifications on RNA Mobility

Using pseudouridine instead of uridine in the IVT mRNA has been critical to therapeutic mRNA efficacy and stability.[235,236] With the reliance on modified mRNA for future therapeutics, it will be essential to characterize if these modifications create a significant impact on the mobility and subsequent characterization of RNA as structurally, Ψ can alter RNA structure by improving base-pairing, base stacking, and contributing to making the backbone more rigid (through a network of hydrogen bonding interactions).[236]

The migration (Figure S7.1a and 7.2a) and effective mobility (Figure 7.5a,b) of modified dsRNA are very similar to those of its non-modified counterpart. While there is a numerical difference between the mobilities of dsRNA and its modified counterpart, the difference is not statistically significant (Figure 7.5a,b). Although a similar direct comparison cannot be made for ssRNA due to sample constraints, the migration profiles (Figure S7.1b and S7.2b) are similar, and the effective mobility of modified RNA fits nicely into the trends seen by its non-modified counterpart (Figure 7.5c). Although the type and extent of modification RNA can undergo will vary depending on the desired application, these findings demonstrate that resulting electrophoretic analysis may not be significantly impacted.

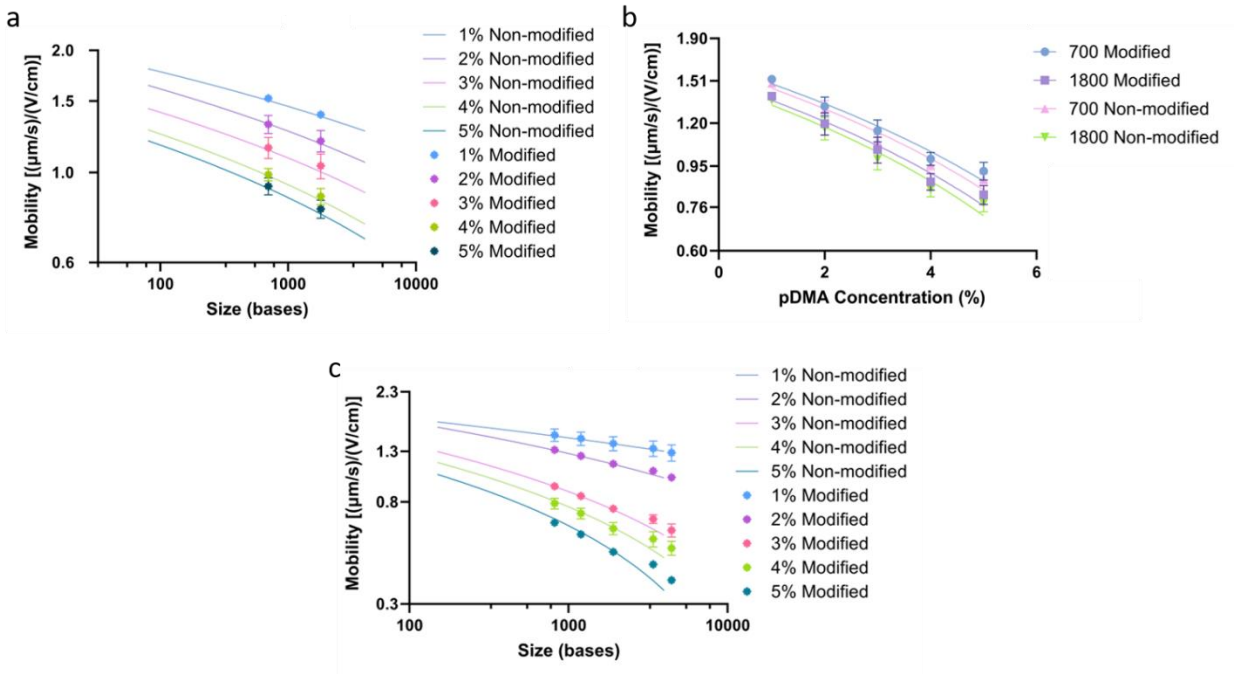

Figure 7.5: Mobility comparison between the non-chemically modified and chemically modified fragments at 1, 2, 3, 4, and 5% pDMA concentrations. (a) Double logarithmic plot of mobility vs. size with the modified dsRNA fragments represented by markers overlayed onto the line of best fit formed from the mobility of its non-modified counterpart. (b) Comparison of modified and non-modified dsRNA mobility of equal sizes in varying PDMA concentrations. (c) Double logarithmic plot of mobility vs. size with the modified ssRNA fragments represented by markers overlayed onto the line of best fit formed from the mobility of its non-modified counterpart.

7.4.4 Machine Learning Modeling for Electrophoretic Mobility Prediction

To enhance the qualitative comparison between our data and established electrophoretic models, we created surrogate models. These models serve as universal function approximators, effectively minimizing the necessity for a large number of physical experiments. To that end, ANNs were used to develop predictive models to solve the following two tasks:

- Task 1: to predict length of the base sequence/number of base pairs (n_b), given the migration time (mt), gel concentration (gc), information about the type of RNA (ssRNA or dsRNA) (tp), and the corresponding persistence length (l_p).
- Task 2: to predict the migration time (mt), given the length of the base sequence/number of base pairs (n_b), gel concentration (gc), information about the type of RNA (ssRNA or dsRNA) (tp), its persistence length (l_p), and the molecular weight of the sequence (mw).

In these two tasks, all the variables except tp have multiple discrete values as tp is used as an identifier with a value of 1 (for ssRNA) or 2 (for dsRNA). To efficiently solve the tasks, we developed a data-driven and physics-informed neural network frameworks. The aim of the physics-driven framework is to augment the learning with the available physical laws along with the available data. For both frameworks, we consider deep neural networks. To that end, we first briefly describe the anatomy of deep neural networks. After that, we discuss various components of the physics-driven approach.

Deep neural networks

Deep neural networks are distinguished from conventional shallow neural networks by the number of hidden layers in the network. In conventional shallow networks, we have one input, output, and hidden layer. On the other hand, in a deep neural network, we have more than one hidden layer;

the intuition is that more hidden layers will be more expressive and provide more accurate results. In this work, we have used a deep, fully connected feed-forward neural network.

Considering that the network consists of L hidden layers, where the 0-th layer denotes the input layer and the L+1-th layer is the output layer, the weighted input, z_i^l into a i-th neuron on a layer, l, is a function of weight, W_{ij}^l and bias, b_j^{l-1} and is represented as:

$$z_i^l = \sigma_{l-1}(\sum_{j=1}^{m_l-1} W_{ij}^l \left(z_i^{l-1}\right) + b_j^l), \tag{6}$$

where $\sigma_{l-1}(\cdot)$ denotes the activation function in the layer l and m_{l-1} are the number of neurons in the layer $l-1$. From the above concepts, the feed-forward algorithm for computing the output, Y^L is expressed as:

$$Y^L = \sigma_L(W^{L+1}z^L + b^L), \tag{7}$$

$$z^L = \sigma_{L-1}(W^L z^{L-1} + b^L), \tag{8}$$

$$z^{L-1} = \sigma_{L-2}(W^{L-1}z^{L-2} + b^{L-1}), \tag{9}$$

$$\vdots$$

$$z^1 = \sigma_0(W^1 x + b^1), \tag{10}$$

where x is the input of the neural network. The above-discussed feed-forward neural networks can be represented in a compressed form as $Y = \mathcal{N}(x;\theta)$, where $\theta = (W, b)$ includes both the weights and biases of the neural network, $\mathcal{N}$. For putting the neural network to use, we need to learn the weights, W_{ij}^l and biases, b_j^{l-1}. Conventionally, this is achieved by first collecting paired data, $D = \{x, Y\}_{i=1}^{N_t}$, and then minimizing a loss function, where N_t denotes the number of training samples.

Common loss functions used in literature include the $\mathcal{L}_1$-loss function and the $\mathcal{L}_2$-loss function, which are defined as:

$$\mathcal{L}_1 = \sum_{i=1}^{n} |Y_i - \mathcal{N}(x_i;\theta)|, \tag{11}$$

$$\mathcal{L}_2 = \sum_{i=1}^{n} |Y_i - \mathcal{N}(x_i;\theta)|^2, \tag{12}$$

where Y_i and $\mathcal{N}(x_i;\theta)$ are the target and corresponding predicted values, respectively, and x_i denotes the sample point.

As discussed previously, one primary bottleneck of data-driven neural networks (both deep and shallow) rests in the fact that a considerable amount of training data is required. In this work, we employ data-augmentation schemes to generate additional labeled datasets given the datasets obtained from the experiments. To that end, we plot the curves of n_b vs. mt for every gc obtained from the experimental data and obtain the equation of the curve using logarithmic regression. For every gc, we generate an additional 20 sample points considering n_b sample points uniformly distributed between 100 and 4000 bases.

For task 1, we aim to learn $\mathcal{N}(\cdot;\theta)$: $[gc, mt, tp, l_p] \rightarrow n_b$. In this scenario, we have four networks, each taking one of the parameters as input to the network. Each network consists of two hidden layers with 32 neurons each. To introduce non-linearity in the network, we have employed a leaky ReLU activation function between the two hidden layers. The connection between the input layer and the first hidden layer as well as the connection between the second hidden layer and the output layer considers linear activation function. The solution field is computed as the inner product of the outputs of the four networks.

For task 1, we have considered the migration time as an input parameter. However, these details might not be available a priori for an unseen case. Therefore, in task 2, we aim to learn the migration time given the other parameters. Hence, we design the network such that $\mathcal{N}(\cdot;\theta)$: $[gc, n_b, mw, tp, l_p] \rightarrow mt$. Similar to the previous task, we design a framework with five deep neural networks consisting of two hidden layers with 32 neurons each. Each network takes as input one of the five quantities in the input space and output a basis function. The solution, mt is written as the inner product of the output embeddings of the five networks. A schematic representation of the framework is presented in Figure 7.6a.

The network parameters are optimized for both tasks using the $\mathcal{L}_2$-loss function and Adam optimizer. The learning rate for the optimizer is 1×10^{-4}. The primary goal for developing ANN-based surrogate models is to generalize well-known to new and unseen data. However, this is a challenging problem. An under-parametrized model with too little capacity cannot learn the problem. In contrast, an over-parametrized model with too much capacity can learn it too well and overfit the training dataset. For our work, we experience overfitting due to the sparse representation of the input space because of the limited availability of labeled data. One popular approach to improve the generalization of deep neural networks is to use regularization during training that keeps the weights of the model small. These techniques not only reduce overfitting, but they can also lead to faster optimization of the model and better overall performance. In the end, we employed $\mathcal{L}_2$ weights regularization with a regularization coefficient of 9×10^{-3}.

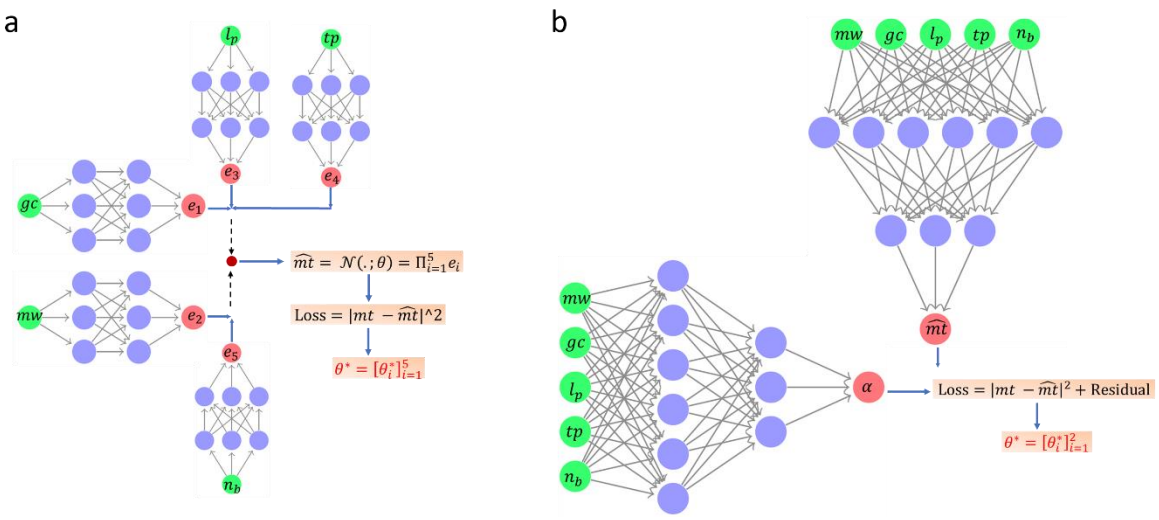

Figure 7.6: Layouts for (a) the data-driven and (b) the PINNS frameworks, where θ^* denotes the collection of optimized parameters of all the networks.

In engineering and biomedical problems, data collection using numerical or physical experiments is often expensive and time-consuming. So, in this work, we also explore physics-informed neural networks (PINNs),[262] where the networks are trained using a modified loss function, which includes the governing equation and the experimental data.

Physics-informed neural networks

This section considers the physics of a problem defined by a generic equation, Equation 7.5. In this work, our goal is to obtain the scalar parameter α along with n_b for task 1 and mt for task 2. For this task, we consider the Biased Reptation model defined as:

$$\frac{\mu}{\mu_0} = \sqrt{\frac{\left(\frac{b}{2lp}\right)^4}{8\frac{n_b^2}{N_l^2}} + \frac{4\epsilon_k^2\left(\frac{b}{2lp}\right)^4}{\left[5+2\alpha\epsilon_k\left(\frac{b}{2lp}\right)^2\right]^2}}\ , \tag{13}$$

where N_l is a constant which is 457 for dsRNA and 14 for ssRNA, μ_0 is the free solution mobility, b is pore size. To that end, we have two deep neural networks consisting of three hidden layers of 8 neurons each and a hyperbolic tan activation function to introduce non-linearity in the PINNs framework (Figure 7.6b). While both the networks take as inputs all the input space parameters discussed above in the data-driven framework, the first network outputs α, and the second network outputs the task-specific quantity of interest (n_b for task 1 and mt for task 2). To construct the loss function, we obtain the residual loss from the governing equation and the $\mathcal{L}_2$ data loss from the experimental data. The standard Adam optimizer minimizes the loss to obtain the optimized network parameters.

An example of the performance of each framework can be observed in Table 7.1, where the predicted values of n_b obtained for ssRNA and dsRNA test samples of Task 1 are compared against the ground truth.

Table 7.1: Comparative results for the predicted lengths of test samples of ssRNA and dsRNA for the PINNs framework for Task 1.

tp		ssRNA				
gc (%)		**1**	**2**	**3**	**4**	**5**
n_b	**500**	486.9	486.9	486.9	486.9	486.9
	1000	988.7	988.7	988.7	988.7	988.7
	2000	1992.3	1992.3	1992.3	1992.3	1992.3
	3000	2995.8	2995.8	2995.8	2995.8	2995.8
	4000	4002.1	4002.1	4002.1	4002.1	4002.1
tp		dsRNA				
n_b	**500**	521.4	521.4	521.4	521.4	521.4
	1000	1008.6	1008.6	1008.6	1008.6	1008.6
	2000	1984.3	1984.3	1984.3	1984.3	1984.3
	3000	2996.8	2996.8	2996.8	2996.8	2996.8
	4000	3996.8	3996.8	3996.8	3996.8	3996.8

Similarly, in Table 2, we highlight the predicted migration times for ssRNA and dsRNA test samples and contrast them to the ground truth.

Table 7.2: Comparative results for the migration time of test samples of ssRNA and dsRNA for the PINNs framework for Task 2. The value in brackets is the ground truth.

tp		ssRNA					dsRNA				
gc (%)		**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
n_b	**500**	24.72 (24.31)	28.55 (27)	38.21 (37.04)	45.92 (42.47)	53.63 (49.12)	23.23 (25.21)	25.83 (28.63)	32.74 (33.21)	38.25 (38.45)	43.75 (41.31)
	1000	27.13 (25.84)	32.57 (29.74)	46.30 (43.33)	57.26 (51.45)	68.21 (61.71)	23.84 (26.51)	27.08 (30.79)	35.25 (35.86)	41.77 (41.95)	48.29 (45.21)
	2000	29.79 (27.86)	36.23 (34.34)	55.22 (55.36)	69.76 (67.60)	84.3 (86.45)	24.57 (28.07)	28.25 (32.94)	37.59 (38.51)	45.06 (45.45)	52.53 (49.12)
	3000	31.21 (28.79)	39.40 (36.21)	60.04 (59.96)	76.52 (73.88)	92.99 (95.57)	24.91 (28.90)	28.87 (34.20)	38.85 (40.06)	46.82 (47.50)	54.77 (51.40)
	4000	32.34 (29.44)	41.27 (37.54)	63.82 (63.23)	81.81 (78.33)	99.80 (102.05)	25.15 (29.53)	29.27 (35.09)	39.65 (41.16)	47.93 (48.96)	56.23 (53.02)

Table 3 summarizes the relative $\mathcal{L}_2$ error (%) computed for five unseen test samples for both the frameworks of ANNs and both tasks. For both tasks, the PINNs model performs better than the data-driven model, and both frameworks perform better at task 1 than task 2.

Table 7.3: Relative $\mathcal{L}_2$ error (%) computed for five unseen test samples for both frameworks of ANNs for both tasks.

	Task 1	Task 2
Data-driven	12.68	18.92
PINNs	0.44	1.1

As expected, given the relatively limited data set, the PINNs model significantly outperforms the data-driven model, as it can compensate for the limited data through the physical equations that govern the experiments. This application of PINNs highlights its potential for electrophoretic analysis, which could far exceed the analysis of single- and double-stranded RNA molecules. Finally, we computed the value of α from the PINNs framework, and report $\alpha = 192.77$. This is essentially an inverse problem, where the network is able to predict the value given the minimization of the residual of the governing equation.

7.5 Conclusion

An in-depth study of the electrokinetic behavior and resulting mobility patterns of dsRNA and ssRNA was conducted, with the desired application of informing future assay protocols and diagnostics. Characterization of RNA will be crucial given the recent advancements in RNA therapeutic development and the subsequent need for quality assurance, primarily due to the potential for RNA degradation during the formulation and storage processes. While current microfluidic chip-based devices for capillary gel electrophoresis face the limitations of providing high-resolution for large RNA, our study was able to characterize RNA mobility for up to around 4,000 bp, which includes the average mRNA length of the human genome as well as the majority of current non-self-amplifying constructs.[276,277] The Ogston-like sieving motion did not describe our data well, indicating limited applicability to RNA of both single-stranded and double-stranded nature above 80 bases. However, predicted transitions between Ogston and Reptation based on

calculations of R_g of nucleic acids and matrix effective pore size were well reflected in our resulting mobility vs. size graphs and may provide guidance when RNA begins to stretch and move through the matrix in a "snake-like" motion. Points of differentiation between dsRNA and ssRNA mobility in entangled polymer solution were identified and may be coupled with other identification methods, as described by our previous work, for complete characterization.

The findings presented in this study serve as an example of how ANNs, particularly PINNs, can be used either in lieu of extensive experiments or to complement limited data sets for assessing the electrophoretic behavior of single- and double-stranded RNA molecules. In this study, using PINNs improved the relative $\mathcal{L}_2$ error from 12.68% using the data-driven model in task 1 to 0.44% and from 18.92% to 1.1% in task 2. This added capability can have significant implications for the biopharmaceutical industry, where machine learning can help streamline the development process, which can involve countless permutations within a given product.

7.6 Acknowledgements

This work was supported by Revvity's research grant to Brown University and by a Brown University Data Science Institute Seed Grant. A.T. is a paid scientific advisor/consultant and lecturer for Revvity.

7.7 Supporting Information

Figure S7.1: Non-chemically modified RNA migration vs. size

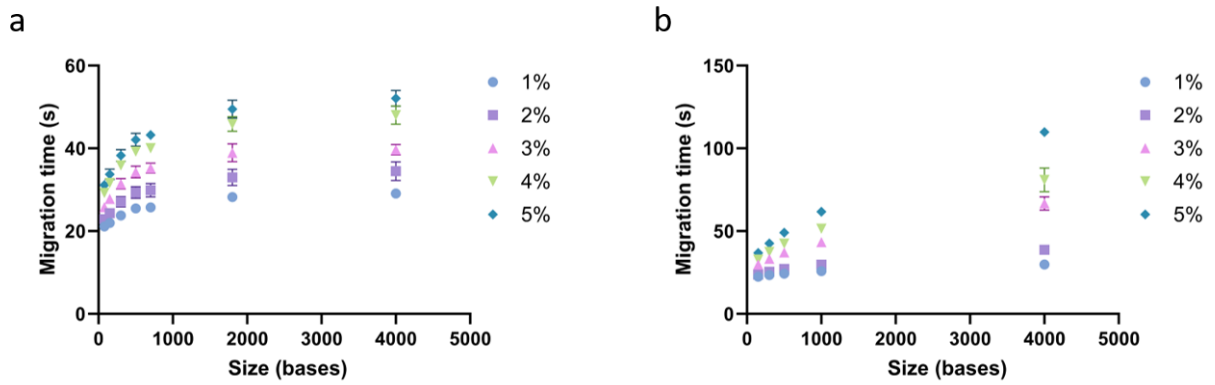

Figure S7.1: Migration vs. size of non-chemically modified (a) dsRNA and (b) ssRNA molecules at 1, 2, 3, 4, and 5% gel concentrations.

Figure S7.2: Chemically modified RNA migration vs. size

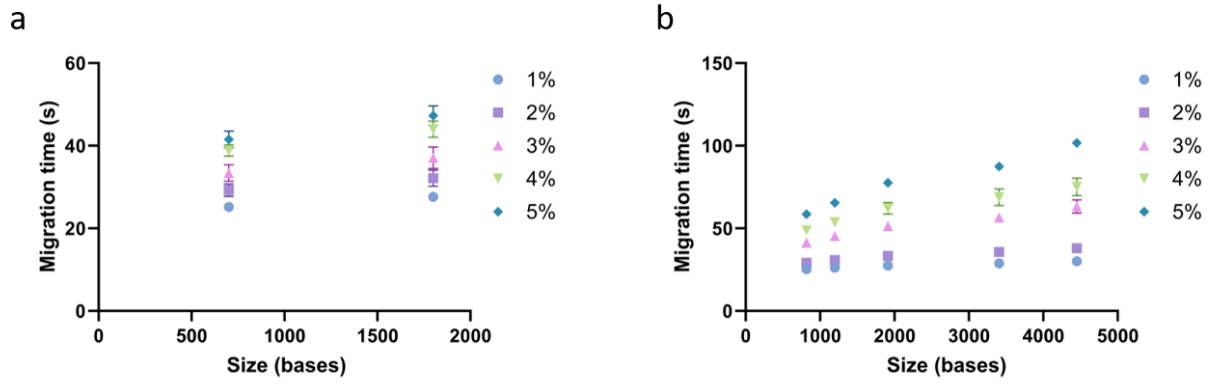

Figure S7.2: Migration vs. size of pseudouridine chemically modified (a) dsRNA and (b) ssRNA molecules at 1, 2, 3, 4, and 5% gel concentrations.

Chapter 8: Assessment of pDNA Isoforms using Microfluidic Electrophoresis

---

Chapter 8 has been submitted as an original research article.

Coll De Peña, A.,[1] Gutterman-Johns, E.,[2] Gautam, G.P.,[3] Rutberg, J.,[1] Ben Frej, M.,[3] Mehta, D.R.,[3] Shah, S.,[3] Tripathi, A.,[1] *Assessment of pDNA Isoforms using Microfluidic Electrophoresis*, **in review**.

[1]Center for Biomedical Engineering, School of Engineering, Brown University, Providence, RI, USA

[2]Department of Molecular Biology, Cell Biology, and Biochemistry, Division of Biology and Medicine, Brown University, Providence, RI, USA

[3]Applied Genomics, Revvity, Hopkinton, MA, USA

8.1 Abstract

The recent rise in nucleic acid-based vaccines and therapies has resulted in an increased demand for plasmid DNA (pDNA). As a result, there is an added pressure to streamline the manufacturing of these vectors, particularly their design and construction, which is currently considered a bottleneck. An important challenge to optimizing pDNA production is the lack of high throughput and rapid analytical methods to support the numerous samples produced during the iterative plasmid construction step and for batch-to-batch purity monitoring. pDNA is generally present as one of three isoforms: supercoiled, linear, or open circular. Depending on the ultimate use, the desired isoform may be supercoiled in the initial stages and for cell transfection or linear in the case of mRNA synthesis. Here, we present a high throughput microfluidic electrophoresis method capable of detecting the three pDNA isoforms and determining the size and concentration of the predominant supercoiled and linear isoforms from 2-7 kb. The limit of detection of the method is 0.1 ng/μL for the supercoiled and linear isoforms and 0.5 ng/μL for the open circular isoform, with a maximum loading capacity of 10-15 ng/μL. The turnaround time is 1 min/sample, and the volume requirement is 10 μL, making the method suitable for process optimization and batch-to-batch analysis. The results presented in this study will enhance the understanding of electrophoretic transport in microscale systems dependent on molecular conformations and potentially aid technological advances in diverse areas relevant to microfluidic devices.

8.2 Introduction

The success of the mRNA-based COVID-19 vaccine, coupled with continued innovation in gene therapies, has led to vastly increased research investment into the potential of nucleic acid in various therapeutic aspects. Plasmid DNA (pDNA) is of significant scientific interest due to its potential uses as a viable vector for gene therapies and DNA vaccinations and as a precursor for the transcription of mRNA therapeutics. Hundreds of biopharmaceutical companies have begun using pDNA specifically for therapeutic development, using it in many currently licensed products.[278] Investigation into gene therapies has continually gained traction year over year, with increased confidence in the future of the field coming from the positive response in the FDA to therapies such as Luxturna and more than 1,500 current clinical trials into gene therapy-related interventions.[279,280] pDNA is a crucial starting point for genetic engineering projects, whether it be the development of recombinant proteins, viral mRNA or DNA vectors, or other advanced biotherapies. However, this growing demand for pDNA in its uses in DNA and RNA-based therapeutics has led to a clear need for standardized pDNA production and quality control processes suitable for clinical applications.[281]

Typically thought of as a circular form of DNA, pDNA and recombinant plasmids have long been synthesized in bacterial hosts, then purified and used in downstream development of therapeutic mRNA or gene therapies. In the synthesis and purification process, however, the plasmids may be damaged and come out of their naturally supercoiled state. The supercoiled isoform is preferred for downstream applications such as gene therapy (i.e., needed for the triple transfection of HEK293 cells for adeno-associated virus synthesis[68,282]) and DNA vaccines, where supercoiled plasmids have shown to have the greatest transfection efficiency.[283,284] In contrast, in the case of mRNA transcription, supercoiled is also the preferred initial conformation, followed by a targeted

linearization before transcription; however, random single-stranded breaks in the sequence will result in an open circle isoform, and proximal double-stranded breaks lead to relatively random linearization of the plasmids,[285] which can affect the sequence of the resulting mRNA. When the plasmids are randomly linearized, the sample is contaminated with a heterogeneous mixture of sequences, which could result in uncontrolled transcripts and potentially dangerous therapeutic side effects. Supercoiled pDNA is the preferred isoform for delivery as a vaccine vector, and pDNA is almost invariably the template in the transcription and production of mRNA vaccines, with a preference for consistent preparation of unvaried formations of pDNA.[286,287] Though plasmids are widely used in basic and applied biology, their design and production remain one of the most time-consuming, labor-intensive processes, and there is no industry-standard method to ensure that the resulting isoforms and sequences are pure enough for therapeutic applications.[288]

The vast scientific and therapeutic potential of pDNA as a vaccine vector and a precursor for gene therapies and mRNA therapeutics has brought about a need for the rapid characterization of the different pDNA isoforms. Currently, the design and construction of plasmids is considered a bottleneck, requiring significant time and labor,[288] further emphasizing the need to develop streamlined, high throughput characterization methods that can support the analytical needs of the numerous samples produced during this process. In particular, it would be beneficial for this method to streamline the optimization of the production conditions and rapidly monitor the presence of the different isoforms in the sample during this optimization and in the final product to ensure the purity of the product. The double-stranded supercoiled isoform of pDNA is widely preferred due to its enhanced efficacy in its uses as a DNA vaccine and gene therapeutic.[289–292] As such, accurate quantification of the possible isoforms is key for use as an mRNA precursor. The critical use of pDNA in these verticals has led to standardized regulation and quality control of

pDNA for its use as a therapeutic, with the FDA requiring > 80% supercoiled pDNA content and various additional governing bodies requiring >90% supercoiled isoform.[283,284] Plasmids used in vaccine trials and as precursors for mRNA vaccines vary in size; however, the upper limit of current clinical relevance appears to be around 10 kb, [285] while the most relevant working range appears to be closer to 4-5 kb.

Several currently validated methods exist for different aspects of pDNA characterization and quantification. Each method has various strengths and weaknesses for specific aspects of pDNA characterization. Numerous electrophoresis methods exist, ranging from standard agarose gel electrophoresis (AGE) to more recently developed sophisticated methods such as Capillary Gel Electrophoresis (CGE).[293,294] These and other gel electrophoresis methods have clear drawbacks, with AGE being a time-consuming and labor-intensive process, CGE requiring specialized machinery, and both having a limited detection range.[295–297] Generally, electrophoretic methods are widely used but often struggle to separate and quantify all three pDNA isoforms, which are of critical importance for downstream applications. Furthermore, a wide range of chromatographic methodology has been developed to be used with an electrophoretic method for greater separation or as an independent method of pDNA QC. These methods include Anion Exchange high-performance liquid Chromatography (AEC), Size Exclusion Chromatography (SEC), and Hydrophobic Interaction Chromatography (HIC).[283,297] Though effective at separation, these methods often require a large sample volume and are better suited for separating pDNA from other biological contaminants than distinguishing between supercoiled, linear, and open circular pDNA isoforms.[283] Furthermore, most chromatographic methods struggle with absolute quantification of the isoforms. Finally, researchers have recently attempted to generate fluorescent-based methods for quantitation of supercoiled pDNA, taking advantage of the unique properties of the supercoiled

isoform. However, this method did not provide accurate measures of linear or open circular pDNA, a key drawback, particularly in using pDNA as a precursor. Auxiliary pDNA QC methods, including a nanodrop analysis and digital PCR, are often used, but both methods are predominantly focused on the quality analysis of just pDNA or pDNA compared to biological contaminants such as proteins, lipids, and DNA fragments.[298] The current best practice for pDNA isoform quantification is Capillary Electrophoresis Laser Induced Fluorescence (CE-LIF); however, the method has a limited detection range in sample concentration and plasmid size. In contrast, AGE is still the preferred method, particularly when working with numerous samples.

With the benefits and limitations of these methods in mind, we have developed a microfluidic, electrophoretic technique to be used in the detection, quantification, and analysis of all isoforms of pDNA for use both in the development of pDNA vaccines and its use as a precursor for other therapeutics. This method has displayed a wide dynamic range of pDNA detection, with accurate isoform separation from 2-7 kb. Furthermore, this high throughput assay can be designed to accommodate 96 or 384 well plates, with easy detection of a large number of samples with minimal labor-intensive preparation and for broad ranges of nucleic acid concentration and plasmid character. Additionally, the method uses only small amounts of sample to accurately measure the entire plasmid batch, providing accurate quantitation without contaminating the therapeutic or precursor load. This method allows for high throughput, accurate assessment of supercoiled, linear, and open circular pDNA isoforms across a wide range of therapeutically relevant sizes and concentrations without major labor or time. This assay has applications in the manufacturing and development of pDNA and mRNA vaccines, particularly when multiple samples need to be prepared and analyzed. Table 8.1 summarizes the current state of the field and the utility of this microfluidic assay in terms of capability and performance.

Table 8.1: Analysis of the current analytical landscape for pDNA isoforms and comparison of their performances. SC stands for supercoiled and OC for open circular.

Method	Isoform Differentiation	Sizing Range	L.O.D.	Volume Requirement	Turnaround Time	High Throughput
CE-LIF[299]	SC, linear, and OC	5-10 kb	5 ng/µL	n/a	20 min/sample	No
AGE[300]	SC, linear, and OC	4.3-10.1 kb	25 ng/µL	60 µL	30-60 minutes	No
CGE[301]	SC, linear, and OC	5-9-15.4 kb	250 ng/µL	> 15 µL	20-30 min/sample	No
HIC[302]	No, no, no	~7 kb	1 ng/µL	500 µL	7 min/sample	No
AEC[303,304]	SC, linear, and OC	< 16 kb	5 ng/µL	50 µL	30 min/sample	No
SEC[305]	Only SC	3-17 kb	12.5 ng/µL	80 µL	10-20 min/sample	No
Microfluidics	SC, linear, and OC	2-7 kb	0.1-0.5 ng/µL	10 µL	1 min/sample	Yes

* Can detect open circular isoforms, but the open circular peaks may co-elute with the linear peaks, leading to poor resolution

8.3 Materials and Methods

8.3.1 Samples and Enzymes

A supercoiled pDNA ladder, with 2.0, 2.5, 3.0, 3.5, 4.0, 5.0, 6.0, 8.0, and 10.0 kb fragments (catalog #B7025, New England Biolabs, Ipswich, MA) was used for the initial assay development, and later for pDNA size determination. A linear DNA ladder with 100, 300, 500, 700, 1,200, 2,000, 3,000, 5,000, 7,000, and 10,000 bp fragments (Revvity, Waltham, MA) was used to fine-tune the method. Three different pDNA vectors (New England Biolabs) - pBR322 (4,361 bp, catalog #N3033), pMAL-c6T (5,247 bp, catalog #N0378), and pTXB1 (6,706 bp, catalog #N6707) – were selected based on their size range and the presence of a BamHI and BsmI restriction site. BamHI-HF (catalog #R3136, New England Biolabs) was used to cleave the plasmids at the BamHI restriction site to linearize the plasmids, while Nb.BsmI (catalog #R0706S, New England Biolabs) induced a single-stranded nick at the BsmI site to generate the open circular isoforms. While BsmI usually generates a double-stranded break, Nb.BsmI was engineered only to induce a single-

stranded nick. Here, it was important to generate the different isoforms in-house to ensure they contained the exact same sequences and could simulate more closely the mobility difference that would be observed in real samples with spontaneous cleavages or nicks.

8.3.2 Sample Preparation for Microfluidic Analysis

The supercoiled samples were diluted in TE buffer, pH 8.0 (ThermoFisher Scientific, Waltham, MA) to the specified concentration. The linear samples were generated using BamHI-HF, following the protocol provided by the manufacturer, by mixing 1 µg of pDNA with 5 µL of 10X rCutSmart Buffer, 1 µL of restriction enzyme, and nuclease-free water to 50 µL incubated at 37°C for 15 min (New England Biolabs), and then diluted in TE buffer (ThermoFisher Scientific) to the specified concentration. Similarly, the open circular samples were generated using Nb.BsmI, following the protocol provided by the manufacturer, by mixing 1 µg of pDNA with 5 µL of 10X NEBuffer, 1 µL of restriction enzyme, and nuclease-free water to 50 µL incubated at 65°C for 60 min and heat-inactivated at 80°C for 20 min, and then diluted in TE buffer (ThermoFisher Scientific) to the specified concentration. Once the samples were generated and combined, when applicable, they were loaded onto a 96- or 384-well plate and transferred onto the LabChip GXII Touch platform (Revvity).

8.3.3 Microfluidic Analysis

A custom microfluidic chip with a metal sipper and reagents for pDNA analysis, detailed below, were provided by Revvity. Before analysis, the microfluidic chip was loaded with a gel-dye mixture. Mixing the fluorescent dye with the gel mixture enables the dynamic labeling of the pDNA molecules as they migrate through the channel. While the dye concentration was kept constant across the experiments, the gel concentration was modified to optimize the separation resolution. A gel diluent was used to decrease the viscosity of the gel-dye mixture while retaining

the conductivity of the mixture. A lower marker was added to one of the wells to be mixed with each sample once loaded onto the chip for alignment. Once the chip was loaded with its reagents, it was transferred with the well plate containing the samples onto the LabChip platform for robotic motion, pressure, and electrokinetic control. Robotic motion was used to move the metal sipper individually over the samples being analyzed, pressure was used to load the sample onto the chip, and electrokinetics was used to generate the movement of the pDNA molecules once in the microfluidic chip. Different electric fields were used to maximize sample separation resolution. After the analysis of each sample, an electropherogram was recorded and analyzed.

8.3.4 Peak Analysis

The peak migration time (t), area, height, and peak width at half-height (w_2) were measured directly. The peak width, w, which is needed to assess the resolution of the assay, was estimated using the following relation:

$$w = 1.7 \times w_2 \tag{8.1}$$

Once the width, w, was estimated, the resolution, R, between any given two peaks was assessed using the following relation:

$$R = \frac{2(t_2 - t_1)}{w_1 + w_2} \tag{8.2}$$

where the subscripts represent the peak number from left to right. The statistical analyses performed in this study, highlighted in the figures, were conducted on GraphPad Prism 9.4.1 with a confidence interval of 95% where * signifies $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and **** $p < 0.0001$.

8.4 Results and Discussion

8.4.1 Development of a Microfluidic Supercoiled pDNA Assay

To develop a robust assay for detecting and characterizing the three main pDNA isoforms, we systematically optimized three key parameters within our system - electrokinetics, gel composition, and sample treatment - to maximize separation resolution and achieve high sensitivity. For the initial development of the assay, we used a supercoiled pDNA ladder (Figure 8.1a) that spans from 2-10 kb. This pDNA ladder was selected and used to optimize the method, as it can also be used downstream to determine the size of the supercoiled isoform present in the samples.

We first conducted initial experiments to narrow the optimal gel concentration range to 0.15% until some individual peaks could be observed using a standard DNA separation script (Figure 8.1b). However, these conditions produced an average resolution of 1.18 ± 0.41 and an average peak height of 17.51 ± 6.64 F.U. (Figure S1a-b), both of which were below the desired ranges. For the purpose of the optimization of the supercoiled ladder assay, the resolution reported here will be the average resolution across adjacent peaks. It is important to note that the fragments in the ladder are not equidistant in length and that it is normal for the resolution to decrease in electrophoretic assay as the fragments get larger; therefore, is not surprising that there will be a large standard deviation associated with the resolutions reported in this study. In addition, while peak area is the parameter relative to concentration, smears are more difficult to detect than sharp peaks, even when their areas are larger; therefore, while we once again expect there to be a wide distribution across peak heights, peak height will be used to assess the sensitivity of the assay. Notwithstanding, the 5 kb peak is expected to be 3× more concentrated than the other fragments, further increasing the wide distribution of peak heights and areas. After optimization of the electrokinetic injection

and separation script for supercoiled pDNA analysis, better-resolved peaks can be observed (Figure 8.1c). These optimizations led to a significant improvement in resolution, with an average of 2.22 ± 0.61, and peak height, with an average of 200.00 ± 107.02 F.U. (Figure S8.1a-b).

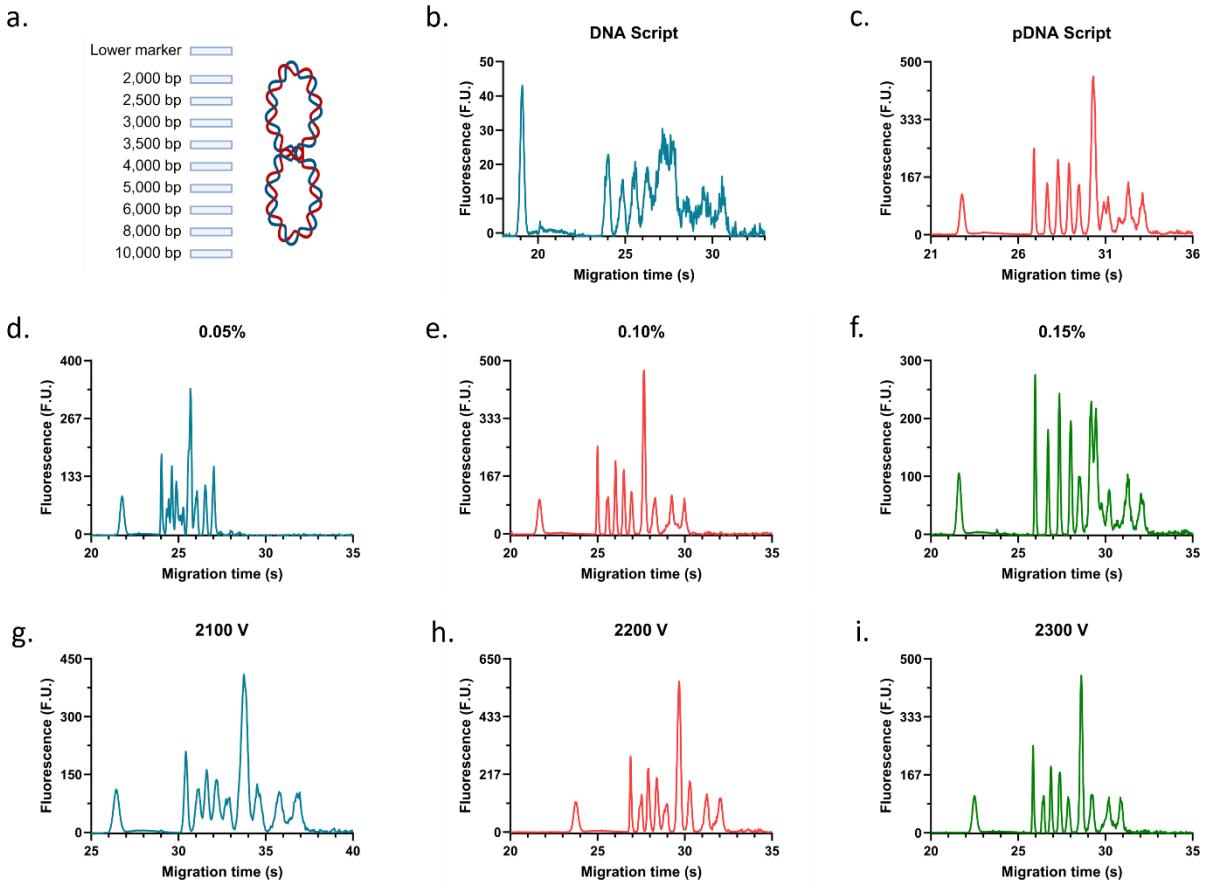

Figure 8.1: Initial method optimization using (a.) a supercoiled ladder with 9 fragments and a system lower marker for alignment. Changes in electropherogram using 0.15% gel using (b.) a standard DNA electrokinetic injection and separation script and (c.) a modified script for supercoiled pDNA analysis. Effect of changes in gel concentration on overall peak shape and separation resolution at (d.) 0.05, (e.) 0.10, and (f.) 0.15% gel. Effect of changes in separation voltage on overall peak shape and separation resolution at (g.) 2100, (h.) 2200, and (i.) 2300V. This figure was created using BioRender.com and GraphPad Prism.

Once the script was roughly optimized, the effect of gel was assessed on the overall electrokinetic mobility of the supercoiled samples (Figure 8.1d-f, Figure S8.1c-d). At 0.05% gel (Figure 8.1d), there was an average resolution of 1.68 ± 0.51 and a peak height of 147.61 ± 81.19 F.U. In the

0.10% gel condition (Figure 8.1e), the average resolution is 2.50 ± 0.49, and the average peak height is 185.68 ± 120.73 F.U. At 0.15% gel (Figure 8.1f), the average resolution is 2.60 ± 1.29, and the average peak height is 163.68 ± 77.80 F.U. Overall, it was observed that while at lower gel concentrations, such as 0.05% (Figure 8.1d), the fragments ≥ 6 kb were better resolved, at the highest gel, 0.15% (Figure 8.1f), the smaller fragments, ≤ 4 kb, were better resolved. While quantitatively, 0.15% may have slightly, although not significantly, higher resolution and peak height values (Figure S8.1c-d), the splitting of the 5 kb peak and significant overlap between the 5 and 6 kb peaks in the 0.15% condition make the 0.10% condition the preferred. Despite the better resolution that can be observed in certain regions of the ladder when splitting takes place, these conditions will not be used as splitting can either prevent the use or decrease the accuracy of using the ladder for size determination downstream.

Once the gel was optimized, since the initial script was optimized for 0.15% gel, the separation voltage was further tuned for 0.10% gel between 2100 and 2300 V (Figure 8.1g-i, Figure S8.1e-f). At 2100 V, the resolution averaged 1.41 ± 0.34 and the peak height 169.02 ± 99.53 F.U., a marked decrease in resolution from previous runs (Figure 8.1g). At 2200 V (Figure 8.1h), the resolution averaged 1.91 ± 0.43 and the peak height 221.53 ± 140.66 F.U., while at 2300 V (Figure 8.1i), the resolution was 2.28 ± 0.51 and the peak height 175.42 ±115.88 F.U., both showing large improvements compared to the 2100 V run. The difference in resolution between our 2100V and 2300V runs was statistically significant, but the results at 2200 V and 2300 V were relatively consistent (Figure S8.1e). With that in mind, a qualitative analysis of the 2300 V condition reveals that while shorter fragments exhibit improved separation, the 8 and 10 kb peaks show greater overlap. This condition also showed a higher likelihood of leading to merged 8 and 10 kb peaks

over multiple runs, which can limit the ability to use the ladder for sizing. As a result, it was ultimately decided to use 0.10% gel with a 2200 V separation with the pDNA script.

8.4.2 Optimization of Separation between Supercoiled and Linear Isoforms

During the initial stages of the study, a method was developed using a supercoiled pDNA ladder to ensure the method could analyze a wide range of supercoiled lengths and because a well-resolved ladder can be used for size determination. However, since the supercoiled and linear isoforms are the two most critical isoforms, it was necessary to validate the method and fine-tune it where needed, using both a supercoiled and a linear ladder. Therefore, here it was essential to ensure both the supercoiled and linear ladders yielded well-resolved peaks for potential use for sizing and to maximize the resolution between supercoiled and linear fragments since, in practice, this is the most relevant separation the method should be capable of achieving. Considering that the supercoiled and linear isoforms typically have similar migration profiles in gel electrophoresis assays, maximizing the resolution between these two isoforms was determined to be the main focus of this section.

Initial tests with the linear ladder using 0.10% gel showed that the longer fragments tended to overlap. Based on the earlier observations, when we changed the gel concentration from 0.05-0.15%, we decided to contrast the separation capabilities of the 0.10% assay, which is run at -2200 V, with a 0.07% run at -2400 V. At 0.07%, the average peak width of our supercoiled ladder was 0.20 ± 0.05 s. In contrast, the average peak width for the linearized ladder was 0.16 ± 0.08 s (Figure 8.2a-b). To determine the theoretical separation capability of the assay at 0.07% gel, the migration times over the length of each ladder were plotted and fitted with logarithmic curves (Figure 8.2c). The fit for the supercoiled ladder curve was $y = 5.1 \times \log(x) + 6.1$, while the linear fit was $y = 2.9 \times \log(x) + 14.6$, with the lines overlapping at a theoretical length of 9.1 kb. In contrast, at

0.10% gel, the average supercoiled ladder peak was 0.52 ± 0.20 s, and the average linear ladder peak was 0.42 ± 25 s (Figure 8.2d-e), which are significantly wider than those observed at 0.07% gel. Interestingly, when their logarithmic fits, $y = 8.9 \times \log(x) + 0.03$ for the supercoiled ladder and $y = 5.1 \times \log(x) + 15.0$ for the linear ladder (Figure 8.2f), the theoretical separation capability was 10.1 kb, 11% higher than at 0.07% gel.

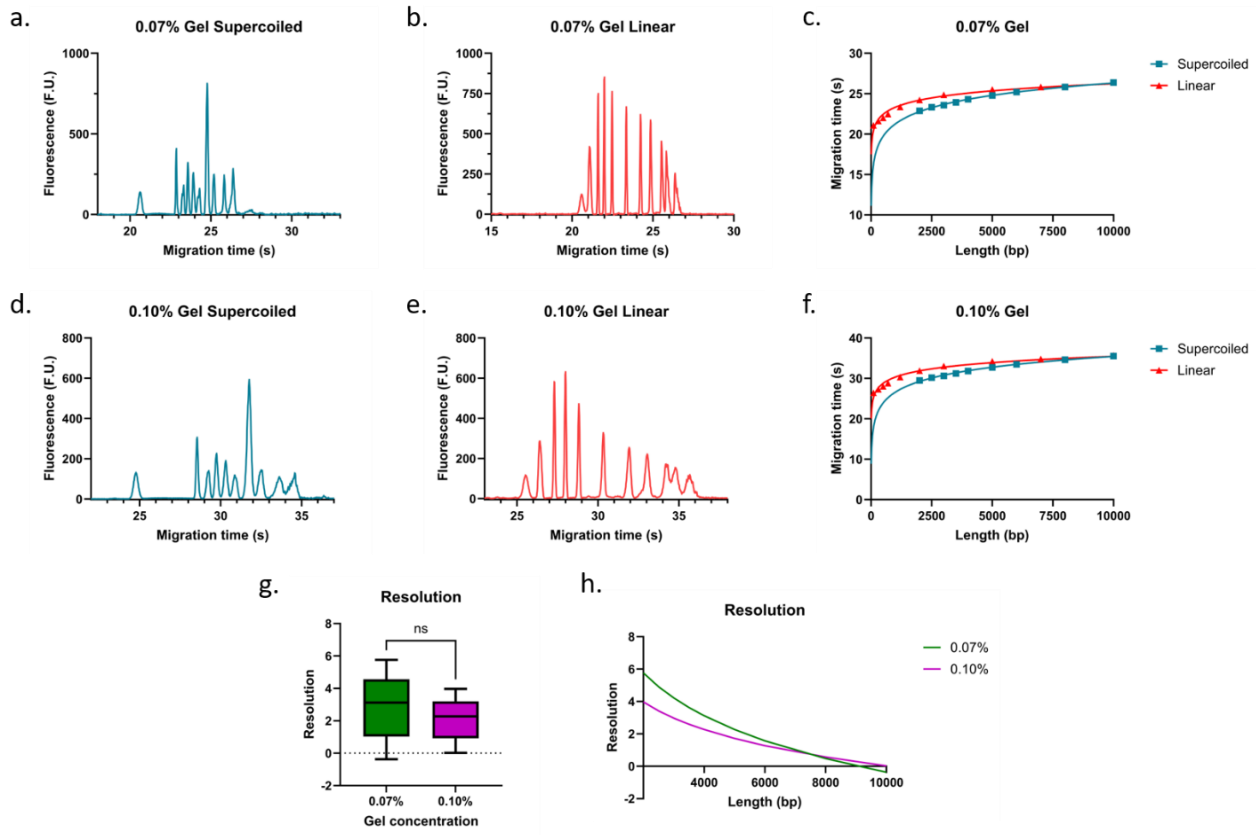

Figure 8.2: Fine-tuning of gel conditions to maximize the resolution between the supercoiled and linear ladders. (a.) Supercoiled and (b.) linear ladders, and (c.) their respective migration profiles at 0.07% gel with 2400 V applied during the separation. (d.) Supercoiled and (e.) linear ladders, and (f.) their respective migration profiles at 0.10% gel with 2200 V applied during the separation. Comparison of (g.) the average resolution and (h.) the resolution over length at 0.07% and 0.10% gel. Here, the statistical significance is represented by * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and **** $p < 0.0001$. This figure was created using GraphPad Prism.

Although the theoretical separation of ladder fragments at 0.10% gel was better, when averaging the resolution of the ladder peak widths (Figure 8.2g), there was no significant difference between

the two gel conditions. Moreover, while the difference in resolution was not statistically significant when all sizes of ladder fragments were averaged, there were apparent differences in resolution when separated by size (Figure 8.2h). When separated by size, as highlighted in Figure 8.2h, it can be observed that the 0.07% gel condition shows better resolution until 7 kb plasmid size. Given that the most critical sizes are 4-5 kb, 0.07% gel was selected as the optimal gel concentration for the remainder of the study, as it has the best supercoiled-linear resolution over the desired size range.

8.4.3 Microfluidic Analysis of pDNA Isoforms

After optimizing the method for separating the supercoiled and linear isoforms, it was important to validate its efficacy in separating low concentrations of the three pDNA isoforms. To that end, three plasmids of varying lengths between 4,361 and 6,706 bp were selected to assess the use of this platform for the sizing and concentration assessment of the three different isoforms. The three plasmids, pBR322, pMAL-c6T, and pTXB1, were purchased as supercoiled samples and, following the procedures described earlier, were digested to generate linear and open circular samples by creating a double- and single-strand nick, respectively, in the supercoiled plasmids. The three isoforms were generated in-house instead of purchased to ensure that the migration profiles of isoforms containing the same sequence could be compared, simulating the composition of real samples where only one sequence should be present across the isoforms.

To assess the performance of the method differentiating the three isoforms, the isoforms were mixed at varying concentrations for each vector, with 3 ng/µL supercoiled, 1ng/µL linear, and 1ng/µL OC. This variation in concentration allows for easy identification of the supercoiled and linear isoforms, which are expected to migrate the closest to each other (Figure 8.3a-c). Interestingly, the supercoiled and linear isoforms display similar peak shapes and magnitudes

relative to their concentration, while the open circular isoform peak is much broader and lower in height (Figure 8.3a-f). The first plasmid tested, pBR322, has a length of 4,361 bp and displayed a clear separation between the three isoforms and an average resolution of 5.15 ± 0.16 between the supercoiled and linear forms (Figure 8.3a,g). As the length of the plasmids increased, there was a slight decrease in resolution between these isoforms, as expected based on the migration time trends observed in Figure 2. For the 5,247 bp plasmid, pMAL-c6T, the average resolution was 3.82 ± 0.79, while for the 6,706 bp plasmid, pTXB1, the average resolution was 2.61 ± 0.00 (Figure 8.3b,c,g). Across the supercoiled-linear resolution of the three vectors, only pBR322 and pTXB1 were statistically different. In contrast, the supercoiled-open circular and linear-open circular resolutions were comparable for the three vectors (Figure 8.3g).

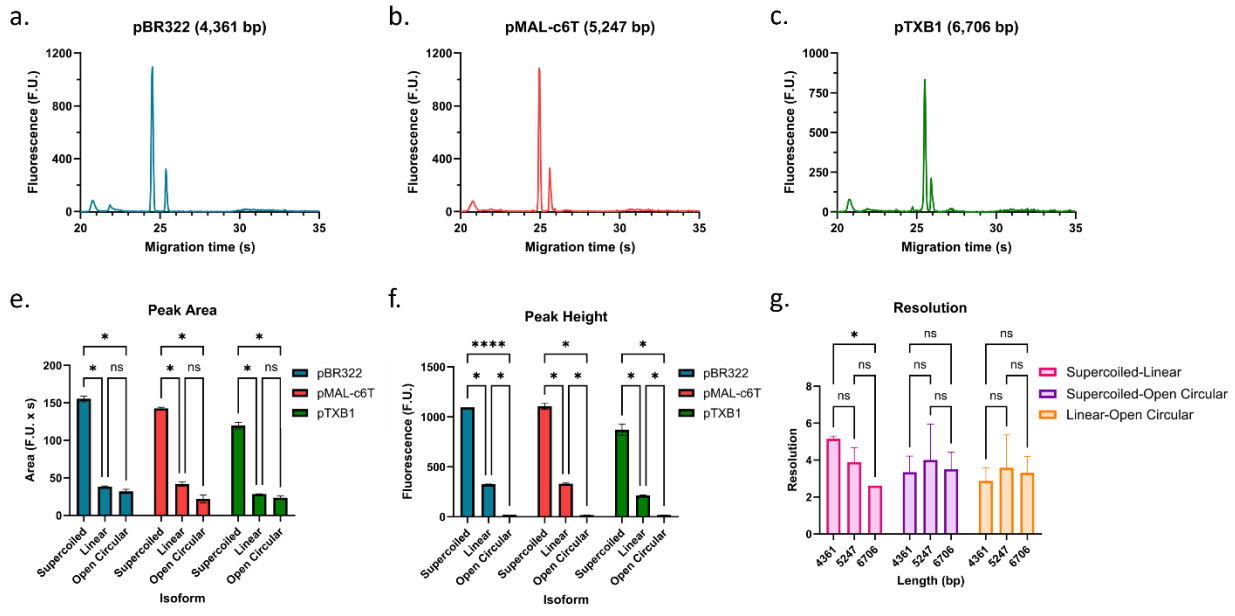

Figure 8.3: Electropherograms of the (a.) pBR322, (b.) pMAL-c6T, and (c.) pTXB1 samples containing 3 ng/µL of supercoiled, 1 ng/µL of linear and 1 ng/µL of open circular isoforms. The concentrations were selected to enable the visual differentiation between the closely migrating supercoiled and linear isoforms. The (e.) peak area, (f.) peak height, and (g.) resolutions of the peaks present in each sample are detailed, and the statistical significance is represented by * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and **** $p < 0.0001$. This figure was created using GraphPad Prism.

Further analysis shows that the migration and mobility profiles of our supercoiled and linear isoforms of each selected plasmid are closely in agreement with the trends observed in our ladder analyses, allowing us to determine the length of the isoforms (Figure 8.4). Each plasmid isoform falls directly on the plotted logarithmic fits of the ladders, confirming the consistency of our assay and indicating the length of a supercoiled or linear plasmid can be determined using a ladder. While the trends of these two isoforms are similar and eventually intersect, the open circular migration and mobility trends are significantly different (Figure 8.4g,h), and due to the lack of an open circular ladder, it is difficult at this time to determine whether the changes in mobility observed are real. Overall, these results prove the potential for pDNA isoform separation and characterization, as relative concentrations of supercoiled and linear isoforms were clearly visible with good resolution between the two peaks, and the migratory and mobility patterns of the isoforms were in agreement with previous trials with a greater variety of plasmid lengths. With further investigation into open circular dynamic trends, this pDNA QC platform shows promise to provide the ability to identify plasmids based on their migration patterns and confirm the relative concentration of each isoform in an unknown sample, a major step in the field.

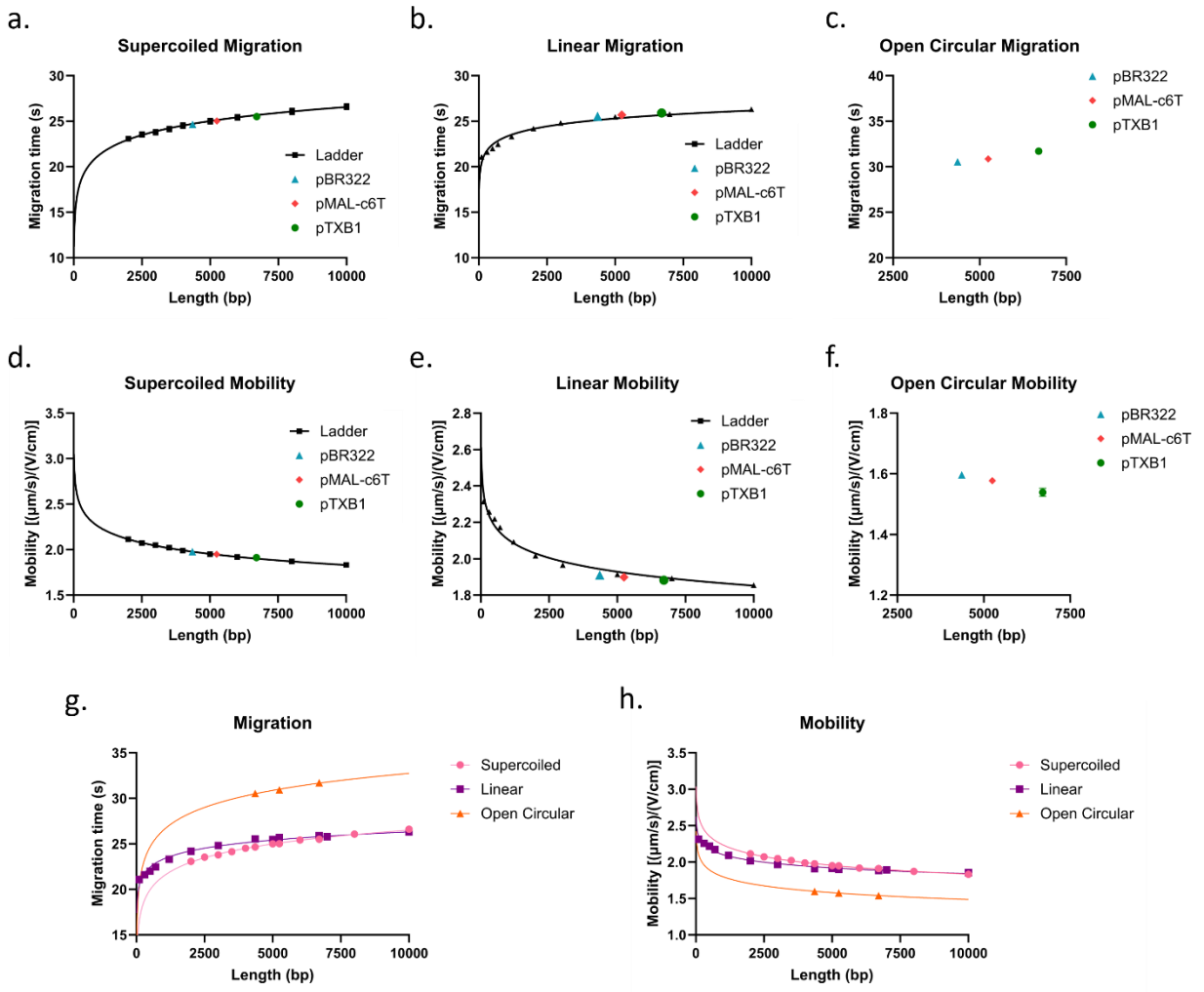

Figure 8.4: (a.) Juxtaposition of the migration profile of the supercoiled ladder fitted with a logarithmic curve and the supercoiled isoforms. (b.) Juxtaposition of the migration profile of the linear ladder fitted with a logarithmic curve and the linear isoforms. (c.) Migration profiles of the open circular isoforms. (d.) Juxtaposition of the mobility profile of the supercoiled ladder fitted with a logarithmic curve and the supercoiled isoforms. (e.) Juxtaposition of the mobility profile of the linear ladder fitted with a logarithmic curve and the linear isoforms. (f.) Mobility profiles of the open circular isoforms. Overlaid (g.) migration and (h.) mobility profiles of the different isoforms. This figure was created using GraphPad Prism.

Once the method was validated with vectors of different lengths, its ability to estimate the concentration of each isoform was tested. To estimate the concentration of unknown samples, the method must show a strong linear relation between peak area and sample concentration. Using the three isoforms of pMAL-c6T at four different concentrations, the correlation between these two parameters was assessed (Figure 8.5). Interestingly, while both the supercoiled and linear isoforms

showed a strong linear relation between peak area and concentration with an R^2 of 0.96, their plots (Figure 8.5d) had significantly different slopes. This finding suggests that the fluorescent dye used in this study has a higher affinity for the supercoiled isoform than for the other two. A significant difference in slope was also observed between the linear and open circular isoforms, with the open circular yielding a significantly smaller signal. Interestingly, the open circular isoform also failed to yield a strong linear relation between the two parameters, with an R^2 of 0.36. This likely suggests that the labeling efficiency of the open circular isoform is very weak and/or that some of the molecules may be getting stuck along the sipper or microfluidic channels. Notwithstanding, the ability to detect the open circular isoform (the peak is not present when the enzyme and buffer are analyzed without the pDNA) still provides a qualitative assessment of the isoform.

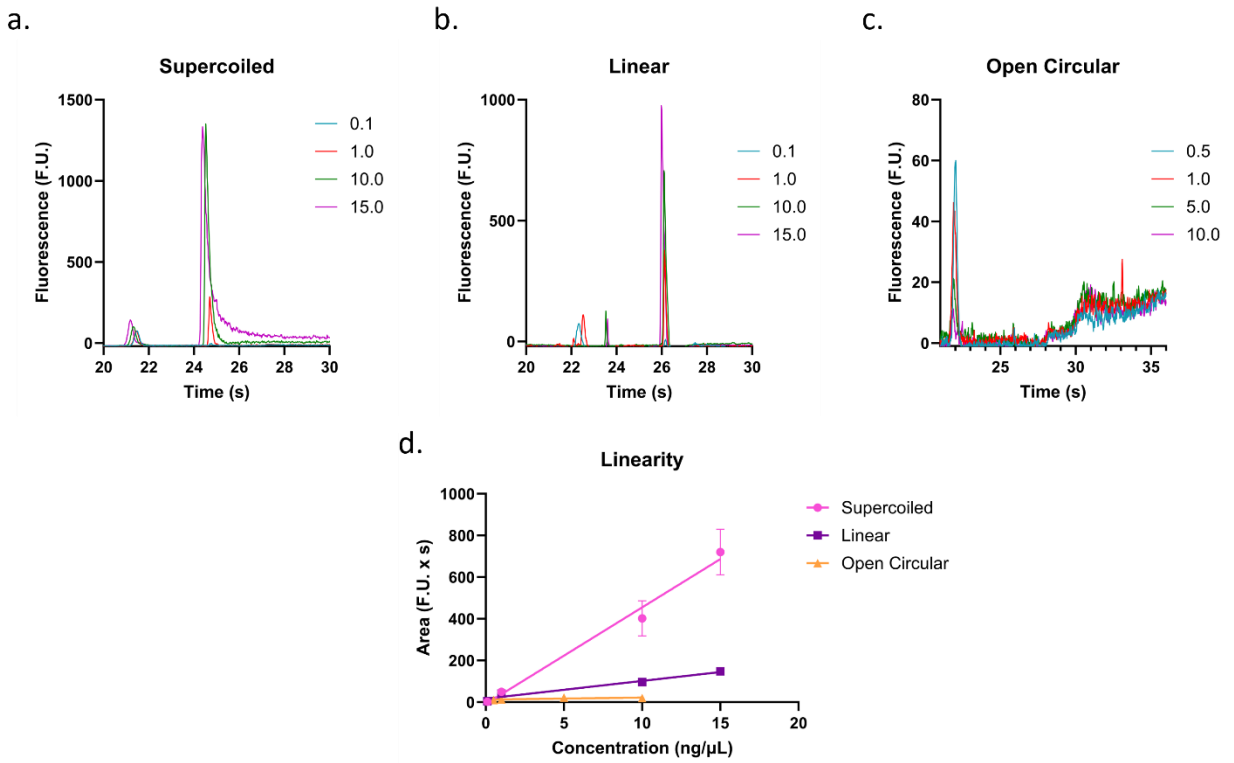

Figure 8.5: Assessment of the linear correlation between isoform peak area and concentration using pMAL-c6T. Electropherograms of the (a.) supercoiled and (b.) linear isoforms at 0.1, 1.0, 10.0, and 15.0 ng/µL and of the (c.) open circular isoform at 0.5, 1.0, 5.0 and 10.0 ng/µL. (d.) Plot of area vs. concentration of the three isoforms where the supercoiled and linear had different slopes but equal R^2 of 0.96. In contrast, the open circular isoform did not show a linear relation between the area and concentration, with its curve yielding an R^2 of 0.36. This figure was created using GraphPad Prism.

8.5 Conclusion

pDNA has become of increasing relevance with the advancements in nucleic acid-based therapeutics. As such, there is an increased need for rapid and robust analytical methods capable of detecting and characterizing its different isoforms, with an increased interest in the supercoiled and linear isoforms. Here, we present a rapid electrophoresis-based analytical method capable of detecting and differentiating the three isoforms of pDNA over a size range of 2-7 kb. It can successfully detect and quantify both the size and concentration of the supercoiled and linear isoforms in the concentration range of 0.1-15.0 ng/µL. However, while it can detect the open

circular isoform at concentrations of 0.5 ng/μL, at this time, without further evaluation and potential method optimization, the assessment of the open circular isoform remains mostly qualitative. Further work is needed to understand better the electrophoretic and fluorescent labeling dynamics of open circular isoforms within microfluidic systems. Notwithstanding, the high throughput compatibility and rapid turnaround time of 1 min/sample of the method make it of particular interest in the context of nucleic acid vaccines and therapeutics as it presents a significant improvement over the currently used analytical tools.

8.6 Acknowledgements

This work was in part supported by Revvity's research grant to Brown University. A.T. is a paid scientific advisor/consultant and lecturer for Revvity.

8.7 Supporting Information

Figure S8.1: Peak height and resolution of the different conditions tested during the initial method optimization

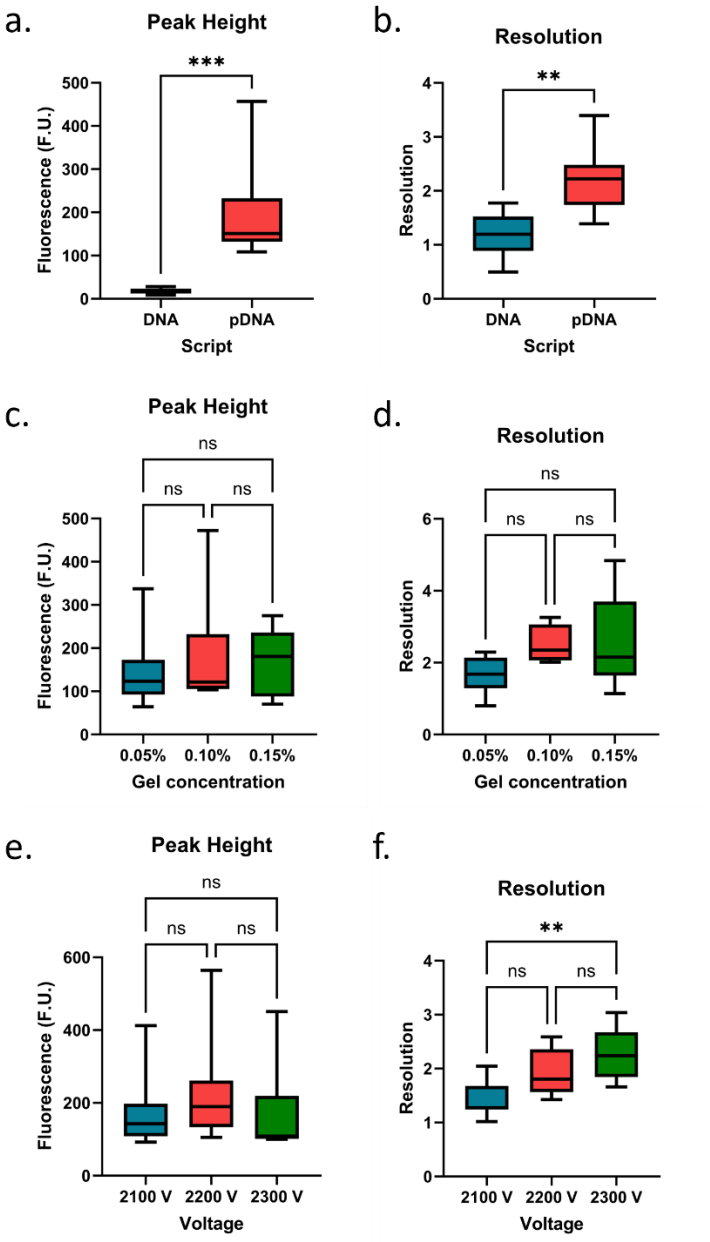

Figure S8.1: (a.) Peak height and (b.) resolution comparison between the DNA and the pDNA scripts. (c.) Peak height and (d.) resolution comparison across 0.05, 0.10 and 0.15% gel conditions. (e.) Peak height and (f.) resolution comparison across 2100, 2200, and 2300 V.

Chapter 9: Conclusions

9.1 Summary

This dissertation focuses on expanding the electrophoretic transport understanding of novel biopharmaceuticals within microfluidic systems. This work provided fundamental insight into the impact of the biochemical and biophysical properties of biopharmaceutical products on their ability to be analyzed and differentiated before and after they were introduced into microscale devices. Given their current clinical relevance in the context of personalized medicine and emerging disease response, adeno-associated viruses, lipid nanoparticles, and nucleic acids were selected as the model systems to explore these interactions. Throughout this work, three main approaches were used to guide this thesis toward these developments: (1) microfluidic electrophoresis sample differentiation, (2) off-chip sample treatment, and (3) quantitative electropherogram data analysis. Conclusions and future directions are described based on these themes.

9.2 Conclusions

9.2.1 Microfluidic Electrophoresis Sample Differentiation

For a microfluidic electrophoresis analytical method to be effective, it must first be capable of separating the samples of interest in space and time, for well-resolved peaks can only be obtained in an electropherogram when samples cross the interrogation window at different time points. To this end, the main parameters that can be changed within the system to differentiate molecules are the plug size, separation voltage, and suspending media. Depending on the desired sensitivity, which depends on the relevant concentration range of the molecules of interest, and on the desired

resolution, which depends on the size and size difference of the molecules of interest, each parameter can be changed to a different extent.

In Chapters 2 and 3, the ssDNA concentration of AAV was very low, and in Chapters 5-7, the dsRNA concentration was very low. Nevertheless, while the fragment lengths varied for both types of molecules, resolution was not as critical since the other molecules present in the sample were expected to have significantly different electrophoretic mobilities at the optimal length. Therefore, in both cases, the plug size, or finite volume of sample, injected into the separation channel was significantly increased to increase the number of molecules of interest that crossed the interrogation probe. However, since increasing the plug decreases the sample plug-to-separation channel length ratio, decreasing the separation potential of the method, this approach would not be suitable in cases where very similar molecules need to be differentiated, as was the case in Chapter 8.

In Chapter 8, the key priority was to maximize the resolution of the separation since the supercoiled and linear isoforms of pDNA have very similar electrophoretic mobilities. In contrast, these samples are usually generated and used at relatively high concentrations, and at the working volume (10 μL), sample scarcity is not a necessary consideration. Therefore, here, the strategy was to decrease the plug size to increase the separation potential of the channel, increase the voltage to limit the time the molecules can diffuse while they reach the interrogation probe to generate narrower peaks and fine-tune the gel viscosity to ensure it is low enough for supercoiled plasmids to separate but high enough for drag to help differentiate the different conformations. Considering the number of parameters that had to be adjusted to achieve the desired resolution, after each parameter was adjusted, the others had to be reevaluated iteratively to ensure they were still appropriate.

9.2.2 Off-Chip Sample Treatment

In some cases, sample complexity and the limited availability of fluorescent probes in the desired operational range make it so that sample labeling cannot be performed dynamically as the molecules migrate through the channel. In other cases, despite the high resolution of the system, the differences between the particles of interest can fall below this threshold. However, diverse biochemical treatments can be integrated into the workflow before microfluidic analysis to label the samples or induce differences based on the fundamental properties of the molecules. Chapters 2, 3, 4, and 6 highlight examples where sample treatments were used.

In Chapters 2 and 3, the full and empty AAV molecules failed to be differentiated intact. However, these two molecules had a major fundamental property that differentiated them: only the full capsids contained a genome. Therefore, it was decided to use two assays to independently assess the number of capsids and ssDNA in the sample, which could be used to determine the percentage of full capsids. For this, to assess multiple attributes simultaneously, the capsids were denatured to enable the assessment of both capsid and individual capsid protein concentration, and a method was developed to digest the capsids to analyze the ssDNA from inside the capsids.

In Chapter 4, the LNPs were first attempted to be differentiated intact. However, these failed to produce any signal due to not interacting directly with a range of fluorescent dyes compatible with the optics range of the system. After exploring different approaches, it was decided to heat the sample with lithium dodecyl sulfate (LDS), known to interact with a fluorescent stain of the desired excitation and emission ranges, until the LDS integrated with the lipids. Once the LDS and LNPs were mixed and heated, they formed new structures that interacted with the fluorescent stain. Upon confirming a linear correlation between the fluorescence yielded by the new structures and sample concentration, the method was used to quantify LNP concentration.

Another example of the use of sample treatment is seen in Chapter 6, where it was used to induce a differential response between dsRNA and mRNA. Here, S1 nuclease was chosen for its selectivity to single-stranded fragments and was used to digest all the mRNA in a dsRNA/mRNA mixture to enable the identification and consequent individual characterization of only the dsRNA contaminants present in the sample. Using this sample treatment also increased the sensitivity of the assay by increasing the maximum loading capacity of the chip.

9.2.3 Quantitative Electropherogram Data Analysis

The last overarching approach guiding this thesis was quantitative electropherogram data analysis. In microfluidic electrophoresis, variations in both migration time and fluorescent magnitudes can be observed across runs, limiting the quantitative potential of the method. However, the integration of controls and standards can significantly mitigate these variations, making microfluidic results highly reproducible. Some standard controls that were integrated across the work presented in this thesis include negative controls to exclude non-sample peaks, markers to align the electropherograms, ladders for size determination, and a control to normalize the signal across runs. An example of an additional approach taken to increase the quantitative potential of the method can be observed in Chapters 2 and 3.

Chapters 2 and 3 aimed to differentiate full from empty AAV capsids by dissecting them into their main components, capsid proteins and ssDNA. The method of independently assessing the capsid and genomic concentrations of AAV and then combining them to assess the percentage of full capsids in the sample is commonly used in the field. However, the direct combination of two independent methods with their respective independent errors led to a compounded error and decreased accuracy. To prevent the generation of a compounded error, here, a reference standard was integrated into each assay, and then a mathematical formulation was developed to combine

the results yielded by both assays by normalizing them to the reference standard. This example highlights how a control and the development of a mathematical formulation were used to analyze the electropherograms more quantitatively and accurately.

9.3 Future Directions

While a wide range of methods were developed for different nucleic acid delivery vehicle products in this thesis, numerous directions could be expanded further to provide appropriate analytical support for the ever-growing and ever-changing biopharmaceutical field.

9.3.1 Translate Methods to Other Viral Platforms and Biopharmaceutical Modalities

The work presented in this thesis focused on characterizing different AAV serotypes; however, variants of AAV are commonly used by biopharmaceutical companies. Therefore, it is also important to validate the applicability and optimize, if needed, the proposed method with more diverse types of AAV samples. In addition, despite AAV being one of the most promising gene therapy platforms, with over 250 clinical trials, lentivirus and adenovirus trials each surpass this number. Due to their important clinical relevance and similar purity analytical needs, it would be valuable to translate this method to these other viral platforms. Interestingly, adenovirus is similar to AAV in that both have non-enveloped protein capsids. However, both adenovirus and the enveloped lentivirus are four to five times bigger than AAV in diameter. These differences make for analytical challenges, but because of the similarities between adenovirus and lentivirus, translating the method to adenovirus first may provide insights that make the more challenging translation to lentivirus less complex. Another potential area to translate this work would be oncolytic viruses, an emerging cancer therapeutics class, as these could also benefit from rapid payload assessment methods. However, this dynamic field has different requirements and viruses of interest, which can pose significant analytical challenges that must be considered.

Lastly, while in this work, great emphasis was placed on nucleic acid delivery vehicles due to their versatility and robustness, microfluidic concepts are also suitable to streamline the characterization and development of other notable biopharmaceutical modalities. Monoclonal antibodies, for instance, are the modality with the most approvals in the market. Nevertheless, while some microfluidic methods with great potential have been developed for their characterization in recent years, their use remains limited, and universal methods have not been established. Working towards expanding the methods and samples compatible with a single analytical modality could significantly maximize the efficiency of the analytical efforts in the industry.

9.3.2 Increase Understanding and Availability of Fluorescent Dyes

While chemists and organic chemists have extensively studied fluorescent dyes, much of this work exists in the chemistry and organic chemistry fields, and a gap exists between those fields and the fields of biomedical science and engineering. In addition, despite the wide range of fluorescent stains and dyes currently available, these are much less abundant when divided by their operational wavelengths, so multiple optics channels or filters must be used. The need for additional optics capabilities was a limitation often encountered during this work that significantly limits the applications of microfluidic analysis to the optical range of the system used. However, considering the high cost associated with optical systems, expanding the offering of fluorescent dyes and stains at different wavelengths may present a better solution that could increase the versatility of the available analytical platforms.

9.3.3 Expand Machine Learning Usage for Electrophoretic Mobility Prediction

In recent years, the role of machine learning in most aspects of life, including biology and medicine, has increased exponentially. In this work, neural networks were explored to complement and possibly decrease the need for experimental data. A key advantage of using machine learning,

as opposed to traditional empirical modeling, is that it enables users to query about conditions that have not been observed. In addition, the recent development of physics-informed neural networks, where physical equations can be fed into the model to decrease the amount of input data needed, has significantly increased the potential for machine learning to support biomedical research, bypassing the need for extensive data sets. Here, the use of this model was explored in terms of the electrophoretic mobility of different RNA molecules to inform the development of analytical methods for RNA vaccines and therapeutics. However, a different molecule-specific data set was used to train the model for each molecule type, even if it can predict the behavior in conditions with which it did not train. In the future, it would be interesting to explore if physics-informed neural networks can be used to predict the electrophoretic behavior of molecules that have not been experimentally tested. However, for this, it is vital to increase the understanding of the governing physical properties responsible for changes in mobility across molecules.

9.4 Contributions of this Dissertation

9.4.1 Contributions to the Field

This dissertation had several contributions to the field, which are summarized in the list below:

- Developed rapid, high-throughput microfluidic analytical methods with a wide range of applications in biomedical research and the biopharmaceutical field.
- Developed a mathematical formulation to determine the percentage of full capsids in AAV samples.
- Provided insight into the interaction between detergents and lipids for fluorescent labeling for LNP analysis.
- Investigated the fundamental properties that affected the electrophoretic mobility of various types of nucleic acid, including ssDNA, pDNA, dsRNA, ssRNA, and mRNA.

- Explored the effect of currents and voltages on the resolution and sensitivity of electrophoretic analyses.
- Investigated fundamental biochemical differences between dsRNA and mRNA in the context of RNA therapeutics and vaccines.

9.4.2 Contributions to the Scientific Community

This thesis has contributed to the scientific community at Brown University and beyond. The contributions to the field are summarized below:

- Mentored eight undergraduates and one high school student for research, innovation, and communication skills.
- Mentored and supervised two undergraduates
- Presented the work at numerous conferences through oral and poster presentations and workshops.
- Explained the work to a lay audience through news articles such as these in Genetic Engineering & Biotechnology News (https://www.genengnews.com/topics/bioprocessing/analyzing-double-stranded-rna-impurities-in-mrna-samples) and in Hoy (https://hoy.com.do/aboga-soluciones-de-biotecnologia-para-patologias).
- Increased the understanding of the molecular transport and electrokinetic behavior of novel biopharmaceuticals within microfluidic systems.
- Explored the benefits of complementing experimental data and findings with machine learning.
- Exploited the biochemical and biophysical properties of molecules and utilized fundamental physics to solve problems using an interdisciplinary approach.

9.4.3 Contributions to the Brown University Community

Coming from the Dominican Republic, where STEM and Biomedical Engineering research are still in their early development, my awareness of biomedical research and the mentorship I received to pursue a career in Biomedical Engineering were limited. Thus, during my Ph.D., I set it as a priority to support and mentor students, particularly women, in pursuing careers in Science and Engineering. As such, a significant focus of my contributions to the Brown University community was surrounding mentorship, diversity, and inclusion. These contributions are listed below:

- Treasurer and member of the Brown University Chapter of the Society for the Advancement of Chicanos/Hispanics and Native Americans in Science.
- Treasurer and organizing committee member for the 2022 New England Regional Society for the Advancement of Chicanos/Hispanics and Native Americans in Science Meeting hosted at Brown University.
- Recruiter for the School of Engineering and the Center for Biomedical Engineering at the 2022 Society for the Advancement of Chicanos/Hispanics and Native Americans in Science National Diversity in STEM Conference.
- Member of the Center for Biomedical Engineering Diversity and Inclusion Action Committee.
- Research mentor of five female undergraduate students and one Ph.D. student in Science and Engineering.
- Peer mentor of three Biomedical Engineering Ph.D. students.
- Co-instructor and teaching assistant for first-year Biology undergraduate student courses.

- Began collaborations with two universities and five industry partners in the United States and abroad.
- Represented Brown University through 14 oral presentations, two posters, and five co-author presentations at regional, national, and international conferences.

References

1. Kesik-Brodacka, M. Progress in biopharmaceutical development. *Biotechnol Appl Biochem* **65**, 306 (2018).

2. Pardi, N., Hogan, M. J. & Weissman, D. Recent advances in mRNA vaccine technology. *Curr Opin Immunol* **65**, 14–20 (2020).

3. Pardi, N., Hogan, M. J., Porter, F. W. & Weissman, D. mRNA vaccines — a new era in vaccinology. *Nature Reviews Drug Discovery 2018 17:4* **17**, 261–279 (2018).

4. Chaudhary, N., Weissman, D. & Whitehead, K. A. mRNA vaccines for infectious diseases: principles, delivery and clinical translation. *Nature Reviews Drug Discovery 2021 20:11* **20**, 817–838 (2021).

5. Roman, G. T. The heights of biopharmaceutical complexity and the current reach of analytical instrumentation. *https://doi.org/10.4155/bio-2017-0118* **9**, 1033–1035 (2017).

6. Whitesides, G. M. The origins and the future of microfluidics. *Nature 2006 442:7101* **442**, 368–373 (2006).

7. Rader, R. A. (Re)defining biopharmaceutical. *Nature Biotechnology 2008 26:7* **26**, 743–751 (2008).

8. Dasani, S., Palanki, R., Menon, P. & Bose, S. K. Biopharmaceuticals. *Handbook for Designing and Conducting Clinical and Translational Surgery* 535–538 (2023) doi:10.1016/B978-0-323-90300-4.00056-2.

9. Walsh, G. & Walsh, E. Biopharmaceutical benchmarks 2022. *Nature Biotechnology 2022 40:12* **40**, 1722–1760 (2022).

10. Mordor Intelligence Inc. Global Biopharmaceutical Market to Grow to Over $600 Billion by 2028. https://blog.marketresearch.com/global-biopharmaceutical-market-to-grow-to-over-600-billion-by-2028 (2023).

11. Mordor Intelligence Inc. Biopharmaceuticals Market - Growth, Trends, Covid-19 Impact, and Forecasts (2023 - 2028). https://www.marketresearch.com/Mordor-Intelligence-LLP-

v4018/Biopharmaceuticals-Growth-Trends-Covid-Impact-33143929/?progid=91829 (2023).

12. Papanikolaou, E. & Bosio, A. The Promise and the Hope of Gene Therapy. *Front Genome Ed* **3**, 618346 (2021).

13. Scheller, E. L. & Krebsbach, P. H. Gene Therapy: Design and Prospects for Craniofacial Regeneration. *J Dent Res* **88**, 585 (2009).

14. Paunovska, K., Loughrey, D. & Dahlman, J. E. Drug delivery systems for RNA therapeutics. *Nature Reviews Genetics 2022 23:5* **23**, 265–280 (2022).

15. Mucker, E. M. *et al.* Lipid Nanoparticle Formulation Increases Efficiency of DNA-Vectored Vaccines/Immunoprophylaxis in Animals Including Transchromosomic Bovines. *Scientific Reports 2020 10:1* **10**, 1–13 (2020).

16. Cui, L. *et al.* Efficient Delivery of DNA Using Lipid Nanoparticles. *Pharmaceutics* **14**, (2022).

17. King, R. A., Rotter, J. I. & Motulsky, A. G. *The Genetic Basis of Common Diseases*. vol. 2 (Oxford University Press, 2002).

18. Qin, S. *et al.* mRNA-based therapeutics: powerful and versatile tools to combat diseases. *Signal Transduction and Targeted Therapy 2022 7:1* **7**, 1–35 (2022).

19. Zu, H. & Gao, D. Non-viral Vectors in Gene Therapy: Recent Development, Challenges, and Prospects. *AAPS Journal* **23**, 1–12 (2021).

20. Alameh, M. G., Weissman, D. & Pardi, N. Messenger RNA-Based Vaccines Against Infectious Diseases. *Curr Top Microbiol Immunol* **440**, 111–145 (2022).

21. Thi, T. T. H. *et al.* Lipid-Based Nanoparticles in the Clinic and Clinical Trials: From Cancer Nanomedicine to COVID-19 Vaccines. *Vaccines (Basel)* **9**, (2021).

22. Patra, J. K. *et al.* Nano based drug delivery systems: recent developments and future prospects. *Journal of Nanobiotechnology 2018 16:1* **16**, 1–33 (2018).

23. Spain, S. G. *et al.* Nanoparticles for Nucleic Acid Delivery. *Comprehensive Biomaterials* **4**, 389–410 (2011).

24. Bulcha, J. T., Wang, Y., Ma, H., Tai, P. W. L. & Gao, G. Viral vector platforms within the gene therapy landscape. *Signal Transduction and Targeted Therapy 2021 6:1* **6**, 1–24 (2021).

25. Sung, Y. K. & Kim, S. W. Recent advances in the development of gene delivery systems. *Biomaterials Research 2019 23:1* **23**, 1–7 (2019).

26. Zaiss, A.-K. *et al.* Differential Activation of Innate Immune Responses by Adenovirus and Adeno-Associated Virus Vectors. *J Virol* **76**, 4580–4590 (2002).

27. Somanathan, S., Breous, E., Bell, P. & Wilson, J. M. AAV Vectors Avoid Inflammatory Signals Necessary to Render Transduced Hepatocyte Targets for Destructive T Cells. *Molecular Therapy* **18**, 977–982 (2010).

28. Müller, D. *et al.* Process intensification in the biopharma industry: Improving efficiency of protein manufacturing processes from development to production scale using synergistic approaches. *Chemical Engineering and Processing - Process Intensification* **171**, 108727 (2022).

29. Carter, L. J. *et al.* Assay Techniques and Test Development for COVID-19 Diagnosis. *ACS Cent Sci* **6**, 591–605 (2020).

30. Udugama, B. *et al.* Diagnosing COVID-19: The Disease and Tools for Detection. *ACS Nano* **14**, 3822–3835 (2020).

31. Yuan, X. *et al.* Current and Perspective Diagnostic Techniques for COVID-19. *ACS Infect Dis* **6**, 1998–2016 (2020).

32. Badihi-Mossberg, M., Buchner, V. & Rishpon, J. Electrochemical Biosensors for Pollutants in the Environment. *Electroanalysis* **19**, 2015–2028 (2007).

33. Li, H., Somerson, J., Xia, F. & Plaxco, K. W. Electrochemical DNA-Based Sensors for Molecular Quality Control: Continuous, Real-Time Melamine Detection in Flowing Whole Milk. *Anal Chem* **90**, 10641–10645 (2018).

34. Ahmed, M. U., Hossain, M. M. & Tamiya, E. Electrochemical Biosensors for Medical and Food Applications. *Electroanalysis* **20**, 616–626 (2008).

35. Kaltashov, I. A. *et al.* Advances and challenges in analytical characterization of biotechnology products: Mass spectrometry-based approaches to study properties and behavior of protein therapeutics. *Biotechnol Adv* **30**, 210–222 (2012).

36. Kaur, H. *et al.* Capillary electrophoresis and the biopharmaceutical industry: Therapeutic protein analysis and characterization. *TrAC Trends in Analytical Chemistry* **144**, 116407 (2021).

37. Hajba, L. & Guttman, A. Recent Advances in the Analysis Full/Empty Capsid Ratio and Genome Integrity of Adeno-associated Virus (AAV) Gene Delivery Vectors. *Curr Mol Med* **20**, 806–813 (2020).

38. Capriotti, A. L., Cavaliere, C. & Piovesana, S. Liposome protein corona characterization as a new approach in nanomedicine. *Anal Bioanal Chem* **411**, 4313–4326 (2019).

39. Subramanian, G. Crude Analytical Assays: PAGE, IEF, Blotting, FTIR, CD, and UV Fluorescence. in *Biopharmaceutical Production Technology* 194–195 (John Wiley & Sons, Incorporated, 2012).

40. Li, T. *et al.* Rapid Determination of Full and Empty Adeno-Associated Virus Capsid Ratio by Capillary Isoelectric Focusing. *Current Molecular Medicine* vol. 20 814–820 Preprint at https://doi.org/http://dx.doi.org/10.2174/1566524020666200915105456 (2020).

41. Rogers, R. S. *et al.* A View on the Importance of 'Multi-Attribute Method' for Measuring Purity of Biopharmaceuticals and Improving Overall Control Strategy. *AAPS J* **20**, (2017).

42. Fernandes, R. P. *et al.* Assessing Multi-Attribute Characterization of Enveloped and Non-Enveloped Viral Particles by Capillary Electrophoresis. *Viruses 2022, Vol. 14, Page 2539* **14**, 2539 (2022).

43. Yang, F. *et al.* Mass spectrometry-based multi-attribute method in protein therapeutics product quality monitoring and quality control. *MAbs* **15**, (2023).

44. Carvalho, S. B. *et al.* Multi attribute method implementation using a High Resolution Mass Spectrometry platform: From sample preparation to batch analysis. *PLoS One* **17**, (2022).

45. Burnham, B. *et al.* Analytical Ultracentrifugation as an Approach to Characterize Recombinant Adeno-Associated Viral Vectors. *Hum Gene Ther Methods* **26**, 228–242 (2015).

46. de Faria, I. J. S., Imler, J. L. & Marques, J. T. Protocol for the analysis of double-stranded RNAs in virus-infected insect cells using anti-dsRNA antibodies. *STAR Protoc* **4**, 102033 (2023).

47. Sanmiguel, J., Gao, G. & Vandenberghe, L. H. Quantitative and Digital Droplet-Based AAV Genome Titration. *Methods Mol Biol* **1950**, 51 (2019).

48. Cristina Poveda, Biter, A. B., Bottazzi, M. E. & Strych, U. Establishing Preferred Product Characterization for the Evaluation of RNA Vaccine Antigens. *Vaccines 2019, Vol. 7, Page 131* **7**, 131 (2019).

49. Silva, T. C., Eppink, M. & Ottens, M. Automation and miniaturization: enabling tools for fast, high-throughput process development in integrated continuous biomanufacturing. *Journal of Chemical Technology & Biotechnology* **97**, 2365–2375 (2022).

50. Tabeling, P. *Introduction to Microfluidics*. (Oxford University Press, 2005).

51. Sackmann, E. K., Fulton, A. L. & Beebe, D. J. The present and future role of microfluidics in biomedical research. *Nature 2014 507:7491* **507**, 181–189 (2014).

52. Lapizco-Encinas, B. H. Chapter 7: Applications of Dielectrophoresis in Microfluidics. in *Microfluidics in detection science: lab-on-a-chip technologies* 192–223 (2014). doi:10.1039/9781849737609-00192.

53. Kirby, B. J. Micro- and Nanoscale Fluid Mechanics: Transport in Microfluidic Devices. *Brian* 512 (2010).

54. Xuan, X. Recent advances in direct current electrokinetic manipulation of particles for microfluidic applications. *Electrophoresis* **40**, 2484–2513 (2019).

55. Gimpel, A. L. *et al.* Analytical methods for process and product characterization of recombinant adeno-associated virus-based gene therapies. *Mol Ther Methods Clin Dev* **20**, 740–754 (2021).

56. Jackson, M., Marks, L., May, G. H. W. & Wilson, J. B. The genetic basis of disease. *Essays Biochem* **62**, 643–723 (2018).

57. Khoury, M. J., Little, J. & Burke, W. *Human Genome Epidemiology: A Scientific Foundation for Using Genetic Information to Improve Health and Prevent Disease*. (Oxford University Press, 2004).

58. Herzog, R. W. & Zolotukhin, S. *A Guide To Human Gene Therapy*. (World Scientific, 2010). doi:10.1142/7406.

59. Mingozzi, F. & High, K. A. Therapeutic in vivo gene transfer for genetic disease using AAV: progress and challenges. *Nat Rev Genet* **12**, 341–355 (2011).

60. Keeler, A. M. & Flotte, T. R. Recombinant Adeno-Associated Virus Gene Therapy in Light of Luxturna (and Zolgensma and Glybera): Where Are We, and How Did We Get Here? *Annu Rev Virol* **6**, 601–621 (2019).

61. Mendell, J. R. *et al.* Current Clinical Applications of In Vivo Gene Therapy with AAVs. *Molecular Therapy* **29**, 464–488 (2021).

62. Agbandje-McKenna, M. & McKenna, R. Production and Purification of Viruses for Structural Studies. in *Structural Virology* P001-371 (The Royal Society of Chemistry, 2011). doi:10.1039/9781849732239.

63. Grieger, J. C. & Samulski, R. J. Adeno-Associated Virus Vectorology, Manufacturing, and Clinical Applications. in *Gene Transfer Vectors for Clinical Application* (ed. Friedmann, T.) vol. 507 229–254 (Academic Press, 2012).

64. Wörner, T. P. *et al.* Adeno-associated virus capsid assembly is divergent and stochastic. *Nature Communications 2021 12:1* **12**, 1–9 (2021).

65. Oyama, H., Ishii, K., Maruno, T., Torisu, T. & Uchiyama, S. Characterization of Adeno-Associated Virus Capsid Proteins with Two Types of VP3-Related Components by Capillary Gel Electrophoresis and Mass Spectrometry. *Hum Gene Ther* **32**, 1403–1416 (2021).

66. Hauswirth, W. W., Lewin, A. S., Zolotukhin, S. & Muzyczka, N. Production and purification of recombinant adeno-associated virus. in *Vertebrate Phototransduction and the Visual Cycle, Part B* vol. 316 743–761 (Academic Press, 2000).

67. Naso, M. F., Tomkowicz, B., Perry, W. L. & Strohl, W. R. Adeno-Associated Virus (AAV) as a Vector for Gene Therapy. *Biodrugs* **31**, 317 (2017).

68. Nguyen, T. N. T. *et al.* Mechanistic model for production of recombinant adeno-associated virus via triple transfection of HEK293 cells. *Mol Ther Methods Clin Dev* **21**, 642–655 (2021).

69. Dong, B., Nakai, H. & Xiao, W. Characterization of Genome Integrity for Oversized Recombinant AAV Vector. *Molecular Therapy* **18**, 87–92 (2010).

70. Khimani, A. H., Thirion, C. & Srivastava, A. AAV Vectors Advance the Frontiers of Gene Therapy. *https://home.liebertpub.com/gen* **42**, 38–40 (2022).

71. Wright, J. F. Quality Control Testing, Characterization and Critical Quality Attributes of Adeno-Associated Virus Vectors Used for Human Gene Therapy. *Biotechnol J* **16**, (2021).

72. Gavin, D. K. FDA Statement Regarding the Use of Adeno-Associated Virus Reference Standard Materials. *Hum Gene Ther Methods* **26**, 3 (2015).

73. Ayuso, E. Manufacturing of recombinant adeno-associated viral vectors: new technologies are welcome. *Mol Ther Methods Clin Dev* **3**, 15049 (2016).

74. Urabe, M. *et al.* Removal of Empty Capsids from Type 1 Adeno-Associated Virus Vector Stocks by Anion-Exchange Chromatography Potentiates Transgene Expression. *Molecular Therapy* **13**, 823–828 (2006).

75. Wright, J. F. Product-Related Impurities in Clinical-Grade Recombinant AAV Vectors: Characterization and Risk Assessment. *Biomedicines 2014, Vol. 2, Pages 80-97* **2**, 80–97 (2014).

76. Penaud-Budloo, M., François, A., Clément, N. & Ayuso, E. Pharmacology of Recombinant Adeno-associated Virus Production. *Mol Ther Methods Clin Dev* **8**, 166–180 (2018).

77. McIntosh, N. L. *et al.* Comprehensive characterization and quantification of adeno associated vectors by size exclusion chromatography and multi angle light scattering. *Sci Rep* **11**, 3012 (2021).

78. Werle, A. K. *et al.* Comparison of analytical techniques to quantitate the capsid content of adeno-associated viral vectors. *Mol Ther Methods Clin Dev* **23**, 254–262 (2021).

79. Burova, E. & Ioffe, E. Chromatographic purification of recombinant adenoviral and adeno-associated viral vectors: methods and implications. *Gene Therapy 2005 12:1* **12**, S5–S17 (2005).

80. Qu, G. *et al.* Separation of adeno-associated virus type 2 empty particles from genome containing vectors by anion-exchange column chromatography. *J Virol Methods* **140**, 183–192 (2007).

81. Wang, C. *et al.* Developing an Anion Exchange Chromatography Assay for Determining Empty and Full Capsid Contents in AAV6.2. *Mol Ther Methods Clin Dev* **15**, 257–263 (2019).

82. Khatwani, S. L., Pavlova, A. & Pirot, Z. Anion-exchange HPLC assay for separation and quantification of empty and full capsids in multiple adeno-associated virus serotypes. *Mol Ther Methods Clin Dev* **21**, 548–558 (2021).

83. Sommer, J. M. *et al.* Quantification of adeno-associated virus particles and empty capsids by optical density measurement. *Molecular Therapy* **7**, 122–128 (2003).

84. Dobnik, D. *et al.* Accurate quantification and characterization of adeno-associated viral vectors. *Front Microbiol* **10**, 1570 (2019).

85. Pierson, E. E., Keifer, David. Z., Asokan, A. & Jarrold, M. F. Resolving Adeno-Associated Viral Particle Diversity With Charge Detection Mass Spectrometry. *Anal Chem* **88**, 6718–6725 (2016).

86. Lock, M., Alvira, M. R. & Wilson, J. M. Analysis of particle content of recombinant adeno-associated virus serotype 8 vectors by ion-exchange chromatography. *Hum Gene Ther Methods* **23**, 56–64 (2012).

87. Wu, D., Hwang, P., Li, T. & Piszczek, G. Rapid characterization of adeno-associated virus (AAV) gene therapy vectors by mass photometry. *Gene Therapy 2021* 1–7 (2022) doi:10.1038/s41434-021-00311-4.

88. Trivedi, P. D. *et al.* Comparison of highly pure rAAV9 vector stocks produced in suspension by PEI transfection or HSV infection reveals striking quantitative and qualitative differences. *Mol Ther Methods Clin Dev* **24**, 154–170 (2022).

89. Horowitz, E. D. *et al.* Biophysical and ultrastructural characterization of adeno-associated virus capsid uncoating and genome release. *J Virol* **87**, 2994–3002 (2013).

90. Venkatakrishnan, B. *et al.* Structure and dynamics of adeno-associated virus serotype 1 VP1-unique N-terminal domain and its role in capsid trafficking. *J Virol* **87**, 4974–4984 (2013).

91. Grieger, J. C. & Samulski, R. J. Adeno-Associated Virus Vectorology, Manufacturing, and Clinical Applications. *Methods Enzymol* **507**, 229–254 (2012).

92. Srivastava, A., Lusby, E. W. & Berns, K. I. Nucleotide sequence and organization of the adeno-associated virus 2 genome. *J Virol* **45**, 555–564 (1983).

93. Li, C. & Samulski, R. J. Engineering adeno-associated virus vectors for gene therapy. *Nat Rev Genet* **21**, 255–272 (2020).

94. Srivastava, A. In vivo tissue-tropism of adeno-associated viral vectors. *Curr Opin Virol* **21**, 75–80 (2016).

95. Zincarelli, C., Soltys, S., Rengo, G. & Rabinowitz, J. E. Analysis of AAV Serotypes 1–9 Mediated Gene Expression and Tropism in Mice After Systemic Injection. *Molecular Therapy* **16**, 1073–1080 (2008).

96. Dobnik, D. *et al.* Accurate Quantification and Characterization of Adeno-Associated Viral Vectors. *Front Microbiol* **10**, (2019).

97. Lock, M., Alvira, M. R., Chen, S.-J. & Wilson, J. M. Absolute Determination of Single-Stranded and Self-Complementary Adeno-Associated Viral Vector Genome Titers by Droplet Digital PCR. *Hum Gene Ther Methods* **25**, 115–125 (2014).

98. Furuta-Hanawa, B., Yamaguchi, T. & Uchida, E. Two-Dimensional Droplet Digital PCR as a Tool for Titration and Integrity Evaluation of Recombinant Adeno-Associated Viral Vectors. *Hum Gene Ther Methods* **30**, 127–136 (2019).

99. Kuck, D., Kern, A. & Kleinschmidt, J. A. Development of AAV serotype-specific ELISAs using novel monoclonal antibodies. *J Virol Methods* **140**, 17–24 (2007).

100. Ayuso, E. *et al.* Manufacturing and Characterization of a Recombinant Adeno-Associated Virus Type 8 Reference Standard Material. *Hum Gene Ther* **25**, 977–987 (2014).

101. Lock, M. *et al.* Characterization of a Recombinant Adeno-Associated Virus Type 2 Reference Standard Material. *Hum Gene Ther* **21**, 1273–1285 (2010).

102. Fu, X. *et al.* Analytical Strategies for Quantification of Adeno-Associated Virus Empty Capsids to Support Process Development. *Hum Gene Ther Methods* **30**, 144–152 (2019).

103. Kohlbrenner, E. & Weber, T. Production and Characterization of Vectors Based on the Cardiotropic AAV Serotype 9. in *Methods in Molecular Biology* vol. 1521 91–107 (2017).

104. Burnham, B. *et al.* Analytical Ultracentrifugation as an Approach to Characterize Recombinant Adeno-Associated Viral Vectors. *Hum Gene Ther Methods* **26**, 228–242 (2015).

105. 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* **110**, 3375–3384 (2021).

106. Gimpel, A. L. *et al.* Analytical methods for process and product characterization of recombinant adeno-associated virus-based gene therapies. *Mol Ther Methods Clin Dev* **20**, 740–754 (2021).

107. Pierson, E. E., Keifer, David. Z., Asokan, A. & Jarrold, M. F. Resolving Adeno-Associated Viral Particle Diversity With Charge Detection Mass Spectrometry. *Anal Chem* **88**, 6718–6725 (2016).

108. McIntosh, N. L. *et al.* Comprehensive characterization and quantification of adeno associated vectors by size exclusion chromatography and multi angle light scattering. *Sci Rep* **11**, 3012 (2021).

109. Coll De Peña, A., Masto, L., Atwood, J. & Tripathi, A. Electrophoresis-Mediated Characterization of Full and Empty Adeno-Associated Virus Capsids. *ACS Omega* **7**, 23457–23466 (2022).

110. Zhang, Y. *et al.* Identification of adeno-associated virus capsid proteins using ZipChip CE/MS. *Anal Biochem* **555**, 22–25 (2018).

111. Katsikis, G. *et al.* Weighing the DNA Content of Adeno-Associated Virus Vectors with Zeptogram Precision Using Nanomechanical Resonators. *Nano Lett* **22**, 1511–1517 (2022).

112. Kotterman, M. A. & Schaffer, D. V. Engineering adeno-associated viruses for clinical gene therapy. *Nature Reviews Genetics 2014 15:7* **15**, 445–451 (2014).

113. Issa, S. S., Shaimardanova, A. A., Solovyeva, V. V. & Rizvanov, A. A. Various AAV Serotypes and Their Applications in Gene Therapy: An Overview. *Cells 2023, Vol. 12, Page 785* **12**, 785 (2023).

114. Münch, R. C. *et al.* Displaying High-affinity Ligands on Adeno-associated Viral Vectors Enables Tumor Cell-specific and Safe Gene Transfer. *Molecular Therapy* **21**, 109–118 (2013).

115. Louis Jeune, V., Joergensen, J. A., Hajjar, R. J. & Weber, T. Pre-existing Anti–Adeno-Associated Virus Antibodies as a Challenge in AAV Gene Therapy. *Hum Gene Ther Methods* **24**, 59–67 (2013).

116. Vandamme, C., Adjali, O. & Mingozzi, F. Unraveling the Complex Story of Immune Responses to AAV Vectors Trial After Trial. *Hum Gene Ther* **28**, 1061–1074 (2017).

117. Bennett, A. *et al.* Thermal Stability as a Determinant of AAV Serotype Identity. *Mol Ther Methods Clin Dev* **6**, 171–182 (2017).

118. Pacouret, S. *et al.* AAV-ID: A Rapid and Robust Assay for Batch-to-Batch Consistency Evaluation of AAV Preparations. *Molecular Therapy* **25**, 1375–1386 (2017).

119. Rayaprolu, V. *et al.* Comparative Analysis of Adeno-Associated Virus Capsid Stability and Dynamics. *J Virol* **87**, 13150–13160 (2013).

120. Srivastava, A. In vivo tissue-tropism of adeno-associated viral vectors. *Curr Opin Virol* **21**, 75 (2016).

121. Addgene: Adeno-associated virus (AAV) Guide. https://www.addgene.org/guides/aav/.

122. ClinicalTrials.gov. https://clinicaltrials.gov/.

123. Srivastava, A. Rationale and strategies for the development of safe and effective optimized AAV vectors for human gene therapy. *Mol Ther Nucleic Acids* **32**, 949–959 (2023).

124. Coll De Peña, A., Masto, L., Atwood, J. & Tripathi, A. Electrophoresis-Mediated Characterization of Full and Empty Adeno-Associated Virus Capsids. *ACS Omega* **7**, 23457–23466 (2022).

125. Johnson, F. B., Ozer, H. L. & Hoggan, M. D. Structural Proteins of Adenovirus-Associated Virus Type 3. *J Virol* **8**, 860 (1971).

126. Snijder, J. *et al.* Defining the Stoichiometry and Cargo Load of Viral and Bacterial Nanoparticles by Orbitrap Mass Spectrometry. *J Am Chem Soc* **136**, 7295 (2014).

127. Girod, A. *et al.* The VP1 capsid protein of adeno-associated virus type 2 is carrying a phospholipase A2 domain required for virus infectivity. *Journal of General Virology* **83**, 973–978 (2002).

128. Venkatakrishnan, B. *et al.* Structure and dynamics of adeno-associated virus serotype 1 VP1-unique N-terminal domain and its role in capsid trafficking. *J Virol* **87**, 4974–4984 (2013).

129. Oyama, H., Ishii, K., Maruno, T., Torisu, T. & Uchiyama, S. Characterization of Adeno-Associated Virus Capsid Proteins with Two Types of VP3-Related Components by Capillary Gel Electrophoresis and Mass Spectrometry. *Hum Gene Ther* **32**, 1403–1416 (2021).

130. Cullis, P. R. & Hope, M. J. Lipid Nanoparticle Systems for Enabling Gene Therapies. *Molecular Therapy* **25**, 1467–1475 (2017).

131. Mitchell, M. J. *et al.* Engineering precision nanoparticles for drug delivery. *Nature Reviews Drug Discovery 2020 20:2* **20**, 101–124 (2020).

132. Kulkarni, J. A., Cullis, P. R. & Van Der Meel, R. Lipid Nanoparticles Enabling Gene Therapies: From Concepts to Clinical Utility. *Nucleic Acid Ther* **28**, 146–157 (2018).

133. Sahin, U., Karikó, K. & Türeci, Ö. mRNA-based therapeutics — developing a new class of drugs. *Nature Reviews Drug Discovery 2014 13:10* **13**, 759–780 (2014).

134. Hou, X., Zaks, T., Langer, R. & Dong, Y. Lipid nanoparticles for mRNA delivery. *Nature Reviews Materials 2021 6:12* **6**, 1078–1094 (2021).

135. Hamilton, A. G., Swingle, K. L. & Mitchell, M. J. Biotechnology: Overcoming biological barriers to nucleic acid delivery using lipid nanoparticles. *PLoS Biol* **21**, e3002105 (2023).

136. Bailey-Hytholt, C. M. *et al.* Intracellular trafficking kinetics of nucleic acid escape from lipid nanoparticles via fluorescence imaging. *Curr Pharm Biotechnol* (2023) doi:10.2174/1389201024666230403094238.

137. Gilleron, J. *et al.* Image-based analysis of lipid nanoparticle–mediated siRNA delivery, intracellular trafficking and endosomal escape. *Nature Biotechnology 2013 31:7* **31**, 638–646 (2013).

138. Schoenmaker, L. *et al.* mRNA-lipid nanoparticle COVID-19 vaccines: Structure and stability. *Int J Pharm* **601**, 120586 (2021).

139. Wilson, B. & Geetha, K. M. Lipid nanoparticles in the development of mRNA vaccines for COVID-19. *J Drug Deliv Sci Technol* **74**, 103553 (2022).

140. Bailey-Hytholt, C. M., Ghosh, P., Dugas, J., Zarraga, I. E. & Bandekar, A. Formulating and Characterizing Lipid Nanoparticles for Gene Delivery using a Microfluidic Mixing Platform. *JoVE* e62226 (2021) doi:doi:10.3791/62226.

141. Hald Albertsen, C. *et al.* The role of lipid components in lipid nanoparticles for vaccines and gene therapy. *Adv Drug Deliv Rev* **188**, 114416 (2022).

142. Tam, Y. Y. C., Chen, S. & Cullis, P. R. Advances in Lipid Nanoparticles for siRNA Delivery. *Pharmaceutics 2013, Vol. 5, Pages 498-507* **5**, 498–507 (2013).

143. Reichmuth, A. M., Oberli, M. A., Jeklenec, A., Langer, R. & Blankschtein, D. mRNA vaccine delivery using lipid nanoparticles. *Ther Deliv* **7**, 319 (2016).

144. Cheng, X. & Lee, R. J. The role of helper lipids in lipid nanoparticles (LNPs) designed for oligonucleotide delivery. *Adv Drug Deliv Rev* **99**, 129–137 (2016).

145. Besin, G. *et al.* Accelerated Blood Clearance of Lipid Nanoparticles Entails a Biphasic Humoral Response of B-1 Followed by B-2 Lymphocytes to Distinct Antigenic Moieties. *Immunohorizons* **3**, 282–293 (2019).

146. Davies, N. *et al.* Functionalized lipid nanoparticles for subcutaneous administration of mRNA to achieve systemic exposures of a therapeutic protein. *Mol Ther Nucleic Acids* **24**, 369–384 (2021).

147. Ball, R. L., Bajaj, P. & Whitehead, K. A. Oral delivery of siRNA lipid nanoparticles: Fate in the GI tract. *Scientific Reports 2018 8:1* **8**, 1–12 (2018).

148. Lokugamage, M. P. *et al.* Optimization of lipid nanoparticles for the delivery of nebulized therapeutic mRNA to the lungs. *Nature Biomedical Engineering 2021 5:9* **5**, 1059–1068 (2021).

149. Wyatt, P. J. Light scattering and the absolute characterization of macromolecules. *Anal Chim Acta* **272**, 1–40 (1993).

150. Wyatt Technology. Dynamic Light Scattering - Size distributions measured quickly and easily. https://www.wyatt.com/solutions/techniques/dynamic-light-scattering-nanoparticle-size.html.

151. Stetefeld, J., McKenna, S. A. & Patel, T. R. Dynamic light scattering: a practical guide and applications in biomedical sciences. *Biophys Rev* **8**, 409–427 (2016).

152. Malvern Panalytical. Electrophoretic Light Scattering (ELS). https://www.malvernpanalytical.com/en/products/technology/light-scattering/electrophoretic-light-scattering.

153. Jia, X. *et al.* Enabling online determination of the size-dependent RNA content of lipid nanoparticle-based RNA formulations. *Journal of Chromatography B* **1186**, 123015 (2021).

154. Takagi, T. Application of low-angle laser light scattering detection in the field of biochemistry : Review of recent progress. *J Chromatogr A* **506**, 409–416 (1990).

155. Wen, J., Arakawa, T. & Philo, J. S. Size-Exclusion Chromatography with On-Line Light-Scattering, Absorbance, and Refractive Index Detectors for Studying Proteins and Their Interactions. *Anal Biochem* **240**, 155–166 (1996).

156. Parot, J., Caputo, F., Mehn, D., Hackley, V. A. & Calzolai, L. Physical characterization of liposomal drug formulations using multi-detector asymmetrical-flow field flow fractionation. *Journal of Controlled Release* **320**, 495–510 (2020).

157. Markova, N. *et al.* Biophysical Characterization of Viral and Lipid-Based Vectors for Vaccines and Therapeutics with Light Scattering and Calorimetric Techniques. *Vaccines (Basel)* **10**, 49 (2021).

158. Brader, M. L. *et al.* Encapsulation state of messenger RNA inside lipid nanoparticles. *Biophys J* **120**, 2766–2770 (2021).

159. Bailey-Hytholt, C. M., Ghosh, P., Dugas, J., Zarraga, I. E. & Bandekar, A. Formulating and Characterizing Lipid Nanoparticles for Gene Delivery using a Microfluidic Mixing Platform. *JoVE (Journal of Visualized Experiments)* **2021**, e62226 (2021).

160. Packer, M., Gyawali, D., Yerabolu, R., Schariter, J. & White, P. A novel mechanism for the loss of mRNA activity in lipid nanoparticle delivery systems. *Nature Communications 2021 12:1* **12**, 1–11 (2021).

161. Chen, C., Chen, C., Li, Y., Gu, R. & Yan, X. Characterization of lipid-based nanomedicines at the single-particle level. *Fundamental Research* **3**, 488–504 (2023).

162. Nag, K. *et al.* DoE-derived continuous and robust process for manufacturing of pharmaceutical-grade wide-range LNPs for RNA-vaccine/drug delivery. *Scientific Reports 2022 12:1* **12**, 1–11 (2022).

163. Mildner, R. *et al.* Improved multidetector asymmetrical-flow field-flow fractionation method for particle sizing and concentration measurements of lipid-based nanocarriers for RNA delivery. *European Journal of Pharmaceutics and Biopharmaceutics* **163**, 252–265 (2021).

164. Kinsey, C. *et al.* Determination of lipid content and stability in lipid nanoparticles using ultra high-performance liquid chromatography in combination with a Corona Charged Aerosol Detector. *Electrophoresis* **43**, 1091–1100 (2022).

165. Schneider, S. & Agilent Technologies Inc. *Analysis of Lipid Nanoparticle Composition*. https://www.agilent.com/cs/library/applications/an-elsd-1260-infinity%20II-prime-bio-lc-1290-infinity%20II-lcsd-5994-4709en-agilent.pdf (2022).

166. Magnusson, L. E., Risley, D. S. & Koropchak, J. A. Aerosol-based detectors for liquid chromatography. *J Chromatogr A* **1421**, 68–81 (2015).

167. Baba, T. *et al.* Dissociation of Biomolecules by an Intense Low-Energy Electron Beam in a High Sensitivity Time-of-Flight Mass Spectrometer. *J Am Soc Mass Spectrom* **32**, 1964–1975 (2021).

168. Kalyanram, P., Puri, A. & Gupta, A. Thermotropic effects of PEGylated lipids on the stability of HPPH-encapsulated lipid nanoparticles (LNP). *J Therm Anal Calorim* **147**, 6337 (2022).

169. Bunjes, H. & Unruh, T. Characterization of lipid nanoparticles by differential scanning calorimetry, X-ray and neutron scattering. *Adv Drug Deliv Rev* **59**, 379–402 (2007).

170. Raffaele, J., Loughney, J. W. & Rustandi, R. R. Development of a microchip capillary electrophoresis method for determination of the purity and integrity of mRNA in lipid nanoparticle vaccines. *Electrophoresis* **43**, 1101–1106 (2022).

171. Karlinsey, J. M. Sample introduction techniques for microchip electrophoresis: A review. *Anal Chim Acta* **725**, 1–13 (2012).

172. Kumar, S. R., Kimchi, E. T., Manjunath, Y., Gajagowni, S. & Stuckel, A. J. RNA cargos in extracellular vesicles derived from blood serum in pancreas associated conditions. *Scientific Reports 2020 10:1* **10**, 1–10 (2020).

173. Peña, A. C. De, Li, N., Vaduva, M., Bwanali, L. & Tripathi, A. A microfluidic electrophoretic dual dynamic staining method for the identification and relative quantitation of dsRNA contaminants in mRNA vaccines. *Analyst* (2023) doi:10.1039/D3AN00281K.

174. Carrasco, M. J. *et al.* Ionization and structural properties of mRNA lipid nanoparticles influence expression in intramuscular and intravascular administration. *Communications Biology 2021 4:1* **4**, 1–15 (2021).

175. Li, Z. *et al.* Acidification-Induced Structure Evolution of Lipid Nanoparticles Correlates with Their In Vitro Gene Transfections. *ACS Nano* **17**, 979–990 (2023).

176. Kamiya, M. *et al.* Stability Study of mRNA-Lipid Nanoparticles Exposed to Various Conditions Based on the Evaluation between Physicochemical Properties and Their Relation with Protein Expression Ability. *Pharmaceutics 2022, Vol. 14, Page 2357* **14**, 2357 (2022).

177. Kumar, V. *et al.* Shielding of Lipid Nanoparticles for siRNA Delivery: Impact on Physicochemical Properties, Cytokine Induction, and Efficacy. *Mol Ther Nucleic Acids* **3**, e210 (2014).

178. Rharbi, Y. & Winnik, M. A. Salt effects on solute exchange in sodium dodecyl sulfate micelles. *J Am Chem Soc* **124**, 2082–2083 (2002).

179. Kumar Singh Tanwar, L., Kumar Banjare, M., Sharma, S. & Ghosh, K. K. Physicochemical studies on the micellization of anionic surfactants in the presence of long alkyl chain ionic liquid. *Chem Phys Lett* **769**, 138399 (2021).

180. Mousli, Y., Brachet, M., Chain, J. L. & Ferey, L. A rapid and quantitative reversed-phase HPLC-DAD/ELSD method for lipids involved in nanoparticle formulations. *J Pharm Biomed Anal* **220**, 115011 (2022).

181. Lichtenberg, D., Ahyayauch, H. & Goñi, F. M. The Mechanism of Detergent Solubilization of Lipid Bilayers. *Biophys J* 289–299 (2013) doi:10.1016/j.bpj.2013.06.007.

182. Coll De Peña, A., White, J. D., Mehta, D. R., Ben Frej, M. & Tripathi, A. Microfluidic AAV Purity Characterization: New Insights into Serotype and Sample Treatment Variability. *ACS Omega* (2024).

183. Quant-iT™ PicoGreen™ dsDNA Assay Kits and dsDNA Reagents. https://www.thermofisher.com/order/catalog/product/P7589?SID=srch-srp-P7589.

184. Eppendorf Research® plus - Mechanical Pipette - Eppendorf US. https://www.eppendorf.com/us-en/eShop-Products/Liquid-Handling/Manual-Pipettes-Dispensers/Eppendorf-Research-plus-p-PF-534798.

185. Kamanzi, A. *et al.* Simultaneous, Single-Particle Measurements of Size and Loading Give Insights into the Structure of Drug-Delivery Nanoparticles. *ACS Nano* **15**, 19244–19255 (2021).

186. Li, S. *et al.* Payload distribution and capacity of mRNA lipid nanoparticles. *Nat Commun* **13**, (2022).

187. Henrickson, A. *et al.* Density Matching Multi-wavelength Analytical Ultracentrifugation to Measure Drug Loading of Lipid Nanoparticle Formulations. *ACS Nano* **15**, 5068–5076 (2021).

188. Staufer, O. *et al.* Bottom-up assembly of biomedical relevant fully synthetic extracellular vesicles. *Sci Adv* **7**, (2021).

189. Rodrigues, C. M. C. & Plotkin, S. A. Impact of Vaccines; Health, Economic and Social Perspectives. *Front Microbiol* **11**, 1526 (2020).

190. Wadhwa, A., Aljabbari, A., Lokras, A., Foged, C. & Thakur, A. Opportunities and Challenges in the Delivery of mRNA-Based Vaccines. *Pharmaceutics* **12**, (2020).

191. Rosa, S. S., Prazeres, D. M. F., Azevedo, A. M. & Marques, M. P. C. mRNA vaccines manufacturing: Challenges and bottlenecks. *Vaccine* **39**, 2190 (2021).

192. Wu, M. Z., Asahara, H., Tzertzinis, G. & Roy, B. Synthesis of low immunogenicity RNA with high-temperature in vitro transcription. *RNA* **26**, 345–360 (2020).

193. de Haro, C., Méndez, R. & Santoyo, J. The eIF-2alpha kinases and the control of protein synthesis. *FASEB J* **10**, 1378–1387 (1996).

194. Minnaert, A. K. *et al.* Strategies for controlling the innate immune activity of conventional and self-amplifying mRNA therapeutics: Getting the message across. *Adv Drug Deliv Rev* **176**, 113900 (2021).

195. McGarry, N. *et al.* Double stranded RNA drives innate immune responses, sickness behavior and cognitive impairment dependent on dsRNA length, IFNAR1 expression and age. *bioRxiv* (2021) doi:10.1101/2021.01.09.426034.

196. McGarry, N. *et al.* Double stranded RNA drives anti-viral innate immune responses, sickness behavior and cognitive dysfunction dependent on dsRNA length, IFNAR1 expression and age. *Brain Behav Immun* **95**, 413 (2021).

197. Martins, R., Queiroz, J. A. & Sousa, F. Ribonucleic acid purification. *J Chromatogr A* **1355**, 1–14 (2014).

198. Karikó, K., Muramatsu, H., Ludwig, J. & Weissman, D. Generating the optimal mRNA for therapy: HPLC purification eliminates immune activation and improves translation of nucleoside-modified, protein-encoding mRNA. *Nucleic Acids Res* **39**, e142–e142 (2011).

199. Baiersdörfer, M. *et al.* A Facile Method for the Removal of dsRNA Contaminant from In Vitro-Transcribed mRNA. *Mol Ther Nucleic Acids* **15**, 26–35 (2019).

200. Son, K.-N., Liang, Z. & Lipton, H. L. Double-Stranded RNA Is Detected by Immunofluorescence Analysis in RNA and DNA Virus Infections, Including Those by Negative-Stranded RNA Viruses. *J Virol* **89**, 9383 (2015).

201. Gabriel, C. J. Detection of double-stranded RNA by immunoblot electrophoresis. *J Virol Methods* **13**, 279–283 (1986).

202. Kartali, T. *et al.* Detection and molecular characterization of novel dsrna viruses related to the totiviridae family in Umbelopsis ramanniana. *Front Cell Infect Microbiol* **9**, 249 (2019).

203. Katri, E. *et al.* Analysis and purification of ssRNA and dsRNA molecules using asymmetrical flow field flow fractionation. *J Chromatogr A* **1683**, 463525 (2022).

204. Dorfman, K. D., King, S. B., Olson, D. W., Thomas, J. D. P. & Tree, D. R. Beyond gel electrophoresis: Microfluidic separations, fluorescence burst analysis, and DNA stretching. *Chem Rev* **113**, 2584 (2013).

205. Abels, J. A., Moreno-Herrero, F., Van Der Heijden, T., Dekker, C. & Dekker, N. H. Single-Molecule Measurements of the Persistence Length of Double-Stranded RNA. *Biophys J* **88**, 2737 (2005).

206. Kebbekus, P., Draper, D. E. & Hagerman, P. Persistence Length of RNA. *Biochemistry* **34**, 4354–4357 (1995).

207. Hyeon, C., Dima, R. I. & Thirumalai, D. Size, shape, and flexibility of RNA structures. *J Chem Phys* **125**, (2006).

208. Bizarro, C. V., Alemany, A. & Ritort, F. Non-specific binding of Na+ and Mg2+ to RNA determined by force spectroscopy methods. *Nucleic Acids Res* **40**, 6922 (2012).

209. Heller, C. Principles of DNA separation with capillary electrophoresis. doi:10.1002/1522-2683.

210. Stellwagen, E. & Stellwagen, N. C. Electrophoretic Mobility of DNA in Solutions of High Ionic Strength. *Biophys J* **118**, 2783–2789 (2020).

211. Heller, C. Separation of double-stranded and single-stranded DNA in polymer solutions: I. Mobility and separation mechanism. *Electrophoresis* **20**, 1962–77 (1999).

212. Yao, Q. *et al.* Differentiating RNA from DNA by a molecular fluorescent probe based on the "door-bolt" mechanism biomaterials. *Biomaterials* **177**, 78–87 (2018).

213. Lipferta, J. *et al.* Double-stranded RNA under force and torque: Similarities to and striking differences from double-stranded DNA. *Proc Natl Acad Sci U S A* **111**, 15408–15413 (2014).

214. To, K. K. W. & Cho, W. C. S. An overview of rational design of mRNA-based therapeutics and vaccines. *Expert Opin Drug Discov* **16**, 1307–1317 (2021).

215. Maruggi, G., Zhang, C., Li, J., Ulmer, J. B. & Yu, D. mRNA as a Transformative Technology for Vaccine Development to Control Infectious Diseases. *Molecular Therapy* **27**, (2019).

216. Feng, R., Patil, S., Zhao, X., Miao, Z. & Qian, A. RNA Therapeutics - Research and Clinical Advancements. *Front Mol Biosci* **8**, 710738 (2021).

217. Barbier, A. J., Jiang, A. Y., Zhang, P., Wooster, R. & Anderson, D. G. The clinical progress of mRNA vaccines and immunotherapies. *Nature Biotechnology 2022 40:6* **40**, 840–854 (2022).

218. Zhang, G., Tang, T., Chen, Y., Huang, X. & Liang, T. mRNA vaccines in disease prevention and treatment. *Signal Transduction and Targeted Therapy 2023 8:1* **8**, 1–30 (2023).

219. Wei, H. H., Zheng, L. & Wang, Z. mRNA therapeutics: New vaccination and beyond. *Fundamental Research* **3**, 749–759 (2023).

220. Triana-Alonso, F. J., Dabrowski, M., Wadzack, J. & Nierhaus, K. H. Self-coded 3'-extension of run-off transcripts produces aberrant products during in vitro transcription with T7 RNA polymerase. *J Biol Chem* **270**, 6298–6307 (1995).

221. Rosa, S. S., Prazeres, D. M. F., Azevedo, A. M. & Marques, M. P. C. mRNA vaccines manufacturing: Challenges and bottlenecks. *Vaccine* **39**, 2190 (2021).

222. Baiersdörfer, M. *et al.* A Facile Method for the Removal of dsRNA Contaminant from In Vitro-Transcribed mRNA. *Mol Ther Nucleic Acids* **15**, 26 (2019).

223. Martínez, J., Lampaya, V., Larraga, A., Magallón, H. & Casabona, D. Purification of linearized template plasmid DNA decreases double-stranded RNA formation during IVT reaction. *Front Mol Biosci* **10**, 1248511 (2023).

224. Zhang, J. *et al.* Recent Advances and Innovations in the Preparation and Purification of In Vitro-Transcribed-mRNA-Based Molecules. *Pharmaceutics 2023, Vol. 15, Page 2182* **15**, 2182 (2023).

225. Piao, X. *et al.* Double-stranded RNA reduction by chaotropic agents during in vitro transcription of messenger RNA. *Mol Ther Nucleic Acids* **29**, 618–624 (2022).

226. Karikó, K., Muramatsu, H., Ludwig, J. & Weissman, D. Generating the optimal mRNA for therapy: HPLC purification eliminates immune activation and improves translation of nucleoside-modified, protein-encoding mRNA. *Nucleic Acids Res* **39**, e142 (2011).

227. Mu, X. & Hur, S. Immunogenicity of In Vitro-Transcribed RNA. *Acc Chem Res* **54**, 4012 (2021).

228. Alexopoulou, L., Holt, A. C., Medzhitov, R. & Flavell, R. A. Recognition of double-stranded RNA and activation of NF-kappaB by Toll-like receptor 3. *Nature* **413**, 732–738 (2001).

229. Mitoma, H. *et al.* The DEAH box RNA helicase DHX33 senses cytosolic RNA and activates the NLRP3 inflammasome. *Immunity* **39**, 123 (2013).

230. McGarry, N. *et al.* Double stranded RNA drives innate immune responses, sickness behavior and cognitive impairment dependent on dsRNA length, IFNAR1 expression and age. *bioRxiv* (2021) doi:10.1101/2021.01.09.426034.

231. Sun, Y., Lu, Z. X., Miller, M., Perroud, T. & Tong, Y. Application of microfluidic chip electrophoresis for high-throughput nucleic acid fluorescence fragment analysis assays. *NAR Genom Bioinform* **5**, (2023).

232. Radpour, R. *et al.* Simultaneous isolation of DNA, RNA, and proteins for genetic, epigenetic, transcriptomic, and proteomic analysis. *J Proteome Res* **8**, 5264–5274 (2009).

233. Slagter-Jäger, J. G., Nicolette, C. A. & Tcherepanova, I. Y. Evaluation of a microfluidics-based platform and slab electrophoresis for determination of size, integrity and quantification of in vitro transcribed RNA used as a component in therapeutic drug manufacturing. *J Pharm Biomed Anal* **70**, 657–663 (2012).

234. Verma, N., Walia, S. & Pandya, A. Micro/nanofluidic devices for DNA/RNA detection and separation. *Prog Mol Biol Transl Sci* **186**, 85–107 (2022).

235. Dolgin, E. The tangled history of mRNA vaccines. *Nature* **597**, 318–324 (2021).

236. Morais, P., Adachi, H. & Yu, Y.-T. The Critical Contribution of Pseudouridine to mRNA COVID-19 Vaccines. *Front Cell Dev Biol* **9**, (2021).

237. Karikó, K. *et al.* Incorporation of Pseudouridine Into mRNA Yields Superior Nonimmunogenic Vector With Increased Translational Capacity and Biological Stability. *Molecular Therapy* **16**, 1833–1840 (2008).

238. Luo, H. T. *et al.* Dissecting the multi-omics atlas of the exosomes released by human lung adenocarcinoma stem-like cells. *npj Genomic Medicine 2021 6:1* **6**, 1–14 (2021).

239. Thakur, B. K. *et al.* Double-stranded DNA in exosomes: a novel biomarker in cancer detection. *Cell Research 2014 24:6* **24**, 766–769 (2014).

240. Ghanam, J. *et al.* DNA in extracellular vesicles: from evolution to its current application in health and disease. *Cell Biosci* **12**, 1–13 (2022).

241. Li, X., Wang, Q. & Wang, R. Roles of Exosome Genomic DNA in Colorectal Cancer. *Front Pharmacol* **13**, (2022).

242. Kirby, T. mRNA vaccine technology for a multivalent flu vaccine. *Lancet Infect Dis* **23**, 157 (2023).

243. Chivukula, S. *et al.* Development of multivalent mRNA vaccine candidates for seasonal or pandemic influenza. *npj Vaccines 2021 6:1* **6**, 1–15 (2021).

244. Arevalo, C. P. *et al.* A multivalent nucleoside-modified mRNA vaccine against all known influenza virus subtypes. *Science (1979)* **378**, 899–904 (2022).

245. Viovy, J.-L. & Duke, T. DNA electrophoresis in polymer solutions: Ogston sieving, reptation and constraint release. *Electrophoresis* **14**, 322–329 (1993).

246. Ogston, A. G. The spaces in a uniform random suspension of fibres. *Transactions of the Faraday Society* **54**, 1754–1757 (1958).

247. Slater, G. W. & Noolandi, J. New Biased-Reptation Model For Charged Polymers. *Phys Rev Lett* **55**, 1579–1582 (1985).

248. Barron, A. E., Sunada, W. M. & Blanch, H. W. Capillary Electrophoresis of DNA in Uncrosslinked Polymer Solutions: Evidence for a N e w Mechanism of DNA Separation. doi:10.1002/(SICI)1097-0290(19961020)52:2.

249. Liu, C., Xu, X., Wang, Q. & Chen, J. Mathematical model for DNA separation by capillary electrophoresis in entangled polymer solutions. *J Chromatogr A* **1142**, 222–230 (2007).

250. Luckey, J. A. & Smith, L. M. A model for the mobility of single-stranded DNA in capillary gel electrophoresis. *Electrophoresis* **14**, 492–501 (1993).

251. Stellwagen, E., Lu, Y. & Stellwagen, N. C. Unified description of electrophoresis and diffusion for DNA and other polyions. *Biochemistry* **42**, 11745–11750 (2003).

252. Liu, C. *et al.* Analysis of small interfering RNA by capillary electrophoresis in hydroxyethylcellulose solutions. *Electrophoresis* **36**, 1651–1657 (2015).

253. Na, H., Kang, B.-H., Ku, J., Kim, Y. & Jeong, K.-H. On-chip Paper Electrophoresis for Ultrafast Screening of Infectious Diseases. *Biochip J* **15**, 305–311 (2021).

254. Yeh, I.-C. & Hummer, G. Diffusion and Electrophoretic Mobility of Single-Stranded RNA from Molecular Dynamics Simulations. *Biophys J* **86**, 681–689 (2004).

255. Ching, T. *et al.* Opportunities and obstacles for deep learning in biology and medicine. *J R Soc Interface* **15**, (2018).

256. Greener, J. G., Kandathil, S. M., Moffat, L. & Jones, D. T. A guide to machine learning for biologists. *Nature Reviews Molecular Cell Biology 2021 23:1* **23**, 40–55 (2021).

257. Lecun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature 2015 521:7553* **521**, 436–444 (2015).

258. McCulloch, W. S. & Pitts, W. A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys* **5**, 115–133 (1943).

259. Raissi, M., Perdikaris, P. & Karniadakis, G. E. Machine learning of linear differential equations using Gaussian processes. *J Comput Phys* **348**, 683–693 (2017).

260. Rudy, S. H., Brunton, S. L., Proctor, J. L. & Kutz, J. N. Data-driven discovery of partial differential equations. *Sci Adv* **3**, (2017).

261. Karniadakis, G. E. *et al.* Physics-informed machine learning. *Nature Reviews Physics 2021 3:6* **3**, 422–440 (2021).

262. Raissi, M., Perdikaris, P. & Karniadakis, G. E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J Comput Phys* **378**, 686–707 (2019).

263. Oliver, G., Simpson, C., Kerby, M. B., Tripathi, A. & Chauhan, A. Electrophoretic migration of proteins in semidilute polymer solutions. *Electrophoresis* **29**, 1152–1163 (2008).

264. Slater, G. W., Noolandi, J. & Eisenberg, A. Radius of gyration of charged reptating chains in electric fields. *Macromolecules* **24**, 6715–6720 (1991).

265. Slater, G. W., Rousseau, J., Noolandi, J., Turmel, C. & Lalande, M. Quantitative analysis of the three regimes of DNA electrophoresis in agarose gels. *Biopolymers* **27**, 509–524 (1988).

266. Abels, J. A., Moreno-Herrero, F., van der Heijden, T., Dekker, C. & Dekker, N. H. Single-molecule measurements of the persistence length of double-stranded RNA. *Biophys J* **88**, 2737–2744 (2005).

267. Chen, H. *et al.* Ionic strength-dependent persistence lengths of single-stranded RNA and DNA. *Proceedings of the National Academy of Sciences* **109**, 799–804 (2012).

268. Herrero-Galán, E. *et al.* Mechanical Identities of RNA and DNA Double Helices Unveiled at the Single-Molecule Level. *J Am Chem Soc* **135**, 122–131 (2013).

269. Todorov, T. I. & Morris, M. D. Comparison of RNA, single-stranded DNA and double-stranded DNA behavior during capillary electrophoresis in semidilute polymer solutions. *Electrophoresis* **23**, 1033–1044 (2002).

270. Lu, T., Klein, L. J., Ha, S. & Rustandi, R. R. High-Resolution capillary electrophoresis separation of large RNA under non-aqueous conditions. *J Chromatogr A* **1618**, 460875 (2020).

271. Chung, M., Kim, D. & Herr, A. E. Polymer sieving matrices in microanalytical electrophoresis. *Analyst* **139**, 5635–5654 (2014).

272. Clos, H. N. & Engelhardt, H. Separations of anionic and cationic synthetic polyelectrolytes by capillary gel electrophoresis. *J Chromatogr A* **802**, 149–157 (1998).

273. Lumpkin, O. J., Déjardin, P. & Zimm, B. H. Theory of gel electrophoresis of DNA. *Biopolymers* **24**, 1573–1593 (1985).

274. Slater, G. W. DNA gel electrophoresis: The reptation model(s). *Electrophoresis* **30**, S181–S187 (2009).

275. Viovy, J.-L. Electrophoresis of DNA and other polyelectrolytes: Physical mechanisms. *Rev Mod Phys* **72**, 813–872 (2000).

276. Bhattacharya, M. *et al.* Bioengineering of Novel Non-Replicating mRNA (NRM) and Self-Amplifying mRNA (SAM) Vaccine Candidates Against SARS-CoV-2 Using Immunoinformatics Approach. *Mol Biotechnol* **64**, 510–525 (2022).

277. NCBI News: Spring 2003 | Human Reference Sequence. https://www.ncbi.nlm.nih.gov/Web/Newsltr/Spring03/human.html.

278. Ohlson, J. Plasmid manufacture is the bottleneck of the genetic medicine revolution. *Drug Discov Today* **25**, 1891–1893 (2020).

279. ClinicalTrials.gov. Search for: Gene Therapy. (2023).

280. Conroy, G. How gene therapy is emerging from its 'dark age'. *Nature* **612**, S24–S26 (2022).

281. Plasmid DNA (pDNA) Production | Sartorius. https://www.sartorius.com/en/applications/biopharmaceutical-manufacturing/regenerative-medicine/gene-therapy/pdna-production.

282. Clément, N. & Grieger, J. C. Manufacturing of recombinant adeno-associated viral vectors for clinical trials. *Mol Ther Methods Clin Dev* **3**, 16002 (2016).

283. Li, H., Bo, H., Wang, J., Shao, H. & Huang, S. Separation of supercoiled from open circular forms of plasmid DNA, and biological activity detection. *Cytotechnology* **63**, 7–12 (2011).

284. Food and Drug Administration. Guidance for Industry: Considerations for Plasmid DNA Vaccines for Infectious Disease Indications. (2007).

285. Schleef, M. & Schmidt, T. Animal-free production of ccc-supercoiled plasmids for research and clinical applications. *J Gene Med* **6**, S45–S53 (2004).

286. Levy, M. S., Lotfian, P., O'Kennedy, R., Lo-Yim, M. Y. & Shamlou, P. A. Quantitation of supercoiled circular content in plasmid DNA solutions using a fluorescence-based method. *Nucleic Acids Res* **28**, e57 (2000).

287. Schlake, T., Thess, A., Fotin-Mleczek, M. & Kallen, K. J. Developing mRNA-vaccine technologies. *RNA Biol* **9**, 1319 (2012).

288. Enghiad, B. *et al.* PlasmidMaker is a versatile, automated, and high throughput end-to-end platform for plasmid construction. *Nature Communications 2022 13:1* **13**, 1–13 (2022).

289. Cupillard, L. *et al.* Impact of plasmid supercoiling on the efficacy of a rabies DNA vaccine to protect cats. *Vaccine* **23**, 1910–1916 (2005).

290. Valente, J. F. A., Sousa, A., Gaspar, V. M., Queiroz, J. A. & Sousa, F. The biological performance of purified supercoiled p53 plasmid DNA in different cancer cell lines. *Process Biochemistry* **75**, 240–249 (2018).

291. Azevedo, G. M. *et al.* Effect of Chromatographic Conditions on Supercoiled Plasmid DNA Stability and Bioactivity. *Applied Sciences 2019, Vol. 9, Page 5170* **9**, 5170 (2019).

292. Hocharoen, L., Noppiboon, S. & Kitsubun, P. Toward QbD Process Understanding on DNA Vaccine Purification Using Design of Experiment. *Front Bioeng Biotechnol* **9**, (2021).

293. Schmeer, M., Buchholz, T. & Schleef, M. Plasmid DNA Manufacturing for Indirect and Direct Clinical Applications. *Hum Gene Ther* **28**, 856–861 (2017).

294. Middaugh, C. R., Evans, R. K., Montgomery, D. L. & Casimiro, D. R. Analysis of plasmid DNA from a pharmaceutical perspective. *J Pharm Sci* **87**, 130–146 (1998).

295. Holovics, H. J., He, Y., Lacher, N. A. & Ruesch, M. N. Capillary gel electrophoresis with laser-induced fluorescence of plasmid DNA in untreated capillary. *Electrophoresis* **31**, 2436–2441 (2010).

296. Masár, M., Hradski, J., Schmid, M. G. & Szucs, R. Advantages and Pitfalls of Capillary Electrophoresis of Pharmaceutical Compounds and Their Enantiomers in Complex Samples: Comparison of Hydrodynamically Opened and Closed Systems. *Int J Mol Sci* **21**, 1–14 (2020).

297. Molloy, M. J. *et al.* Effective and robust plasmid topology analysis and the subsequent characterization of the plasmid isoforms thereby observed. *Nucleic Acids Res* **32**, e129–e129 (2004).

298. Dong, L., Yoo, H. B., Wang, J. & Park, S. R. Accurate quantification of supercoiled DNA by digital PCR. *Scientific Reports 2016 6:1* **6**, 1–9 (2016).

299. Luo, J., Li, T. & Malik, M. Simple, fast and robust method for plasmid purity testing and degradation monitoring. *Sciex*.

300. Shubsda, M. F., Goodisman, J. & Dabrowiak, J. C. Quantitation of ethidium-stained closed circular DNA in agarose gels. *J Biochem Biophys Methods* **34**, 73–79 (1997).

301. Wang, M. *et al.* A platform method for plasmid isoforms analysis by capillary gel electrophoresis. *Electrophoresis* **43**, 1174–1182 (2022).

302. Diogo, M. M., Queiroz, J. A. & Prazeres, D. M. F. Assessment of purity and quantification of plasmid DNA in process solutions using high-performance hydrophobic interaction chromatography. *J Chromatogr A* **998**, 109–117 (2003).

303. Pavlin, N. *et al.* Analytical separation of plasmid DNA isoforms using anion exchanging chromatographic monoliths with 6 µm channels. *Electrophoresis* (2023) doi:10.1002/ELPS.202300031.

304. Kralj, Š. *et al.* Effect of plasmid DNA isoforms on preparative anion exchange chromatography. *Electrophoresis* (2023) doi:10.1002/ELPS.202300035.

305. Latulippe, D. R. & Zydney, A. L. Size exclusion chromatography of plasmid DNA isoforms. *J Chromatogr A* **1216**, 6295–6302 (2009).