BIOPHYSICAL CHARACTERIZATION OF NON-CANONICAL VIRAL NUCLEIC ACID INTERACTIONS WITH HUMAN PROTEINS

DANIELLE LUNA GEMMILL
Bachelor of Science, University of Lethbridge, 2018

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

DOCTOR OF PHILOSOPHY
in
BIOMOLECULAR SCIENCE – BIOPHYSICS

Department of Chemistry and Biochemistry
University of Lethbridge
Lethbridge, Alberta, Canada

BIOPHYSICAL CHARACTERIZATION OF NON-CANONICAL VIRAL NUCLEIC ACID INTERACTIONS WITH HUMAN PROTEINS

Danielle Luna Gemmill

Date of Defense: April 1, 2025

Dr. Trushar Patel Thesis Supervisor	Associate Professor University of Lethbridge	Ph.D.
Dr. Tony Russell Committee Member	Associate Professor University of Lethbridge	Ph.D.
Dr. Wade Abbott Committee Member	Adjunct Professor University of Lethbridge	Ph.D.
Dr. Jason Kindrachuk External Examiner	Associate Professor University of Manitoba	Ph.D.
Dr. Michael Gerken Chair, Thesis Examination Committee	Professor University of Lethbridge	Ph.D.

Dedication

This dissertation is dedicated to Gabe, my children, and all individuals who stand in solidarity with me in supporting my pursuits in research. Thank you from the deepest core of my being.

Abstract

Viruses are a replicative, paradoxical genetic element that are critical to our survival, but are also a significant burden to public health. Viruses that cause disease in humans continuously evolve to avoid counteroffensive measures from our immune system. Both DNA and RNA viruses have developed various methods to optimize their replication and spread to a host, such as genomic mutations that renders host antibodies ineffective, or by possessing structure that confers an advantage for propagation of virions. Moreover, the structures that viruses possess within their genomes alone provide the virus with the ability to hijack host proteins for replicative and propagation purposes by hijacking and suppression of host cellular functions, altogether benefitting the viral lifecycle. This thesis focuses on the identification of the viral nucleic acid structures that DNA and RNA viruses possess that allow them to replicate and avoid host defenses, and as well, identify host proteins being hijacked by these structures. Specifically, I characterize conserved regions of Zika virus known as the terminal regions that regulate transcription and translation of the viral genome and perform immunoprecipitation pulldown assay to identify host binding partners interacting with them that are critical for viral replication. I also elucidated a non-canonical structure (G-quadruplex) in these terminal regions that interacts with host proteins that could be utilized as a potential anti-viral target, and further, show that G-quadruplexes are also present in DNA viruses such as Mpox virus. I show that TMPyP4, a G-quadruplex binding small molecule, interacts with the Mpox G-quadruplexes and reduces viral protein production.

Contribution of Authors

This thesis is a collection of my doctoral research studies in the field of biophysical virology, conducted at the University of Lethbridge under the supervision of Dr. Trushar Patel. The work presented herein focuses on the structural studies of viral DNA and RNA, and their interactions with human proteins. Over the course of my studies, I have employed a multidisciplinary approach, combining techniques from molecular biology, biochemistry, biophysics, chemistry, and cellular biology. This research not only contributes to our fundamental understanding of biophysical virology but lays a foundation for the development of novel antiviral strategies and countermeasures. Portions of this thesis have been published in scientific literature in the Journal of Biochemistry and Cell Biology, as well as the Journal of Medical Virology. These publications are reflected in Chapters 3, 6, and 8, and the findings presented elucidate viral replication and build a foundation to explore antiviral strategies to target viruses in general. Chapter 3, Danielle Gemmill wrote the entire manuscript, with Simmone D'souza, Vanessa Meier-Stephenson and Trushar R. Patel contributing for editing and giving suggestions on the manuscript writing. Chapter 6, Danielle Gemmill wrote the entire manuscript, and made all figures, and performed all experiments and data analysis except for the locARNA alignment and figure, and the MST DDX17/DDX3X interaction with the 3'TR. These experiments and figures were performed and made by Michael T. Wolfinger and Corey R. Nelson, respectively. Maulik D. Badmalia, Higor S. Pereira assisted in editing the manuscript, and providing insight into some of the experimentation. Liam Kerr assisted with the fluorimetry experiments. Finally, Chapter 8 Danielle Gemmill performed all G4 purification and folding preparation, CD spectroscopy, preparation for SAXS, Thioflavin T assay, DHX36 expression and purification, DHX36 and TMPyP4 binding affinity experiments with MST, and performed the DHX36 and TMPyP4 competing assays alongside Higor Sette Pereira. I additionally co-wrote the introduction to the manuscript, as well as the CD spectroscopy and MST sections. Higor Sette Pereira wrote the rest of the manuscript, as well as performed the SAXS data analysis and structure determination, and performed alongside M. Quadir Siddiqui, the mammalian cell culture experimentation with the MKPX A27L and A50R genes and RFP reporter system. Gunjan Vasudeva assisted in genomic analysis, as well as assisting in editing the manuscript, and Trushar R. Patel provided insight and oversaw the project. Chapter 5 on Nucleolin was 85% contribution of Danielle Gemmill, with Dr. Higor Pereira assisting with protein expression optimization, and Dr. Laurent Chatel-Chaix's laboratory assisting with cell culture studies. All other chapters in my thesis are 100% contribution from Danielle Gemmill.

I am grateful for the support and mentorship of Dr. Patel, my laboratory team, and the broader scientific community that have altogether fostered my growth as a researcher. It's my sincere hope that this thesis will contribute meaningfully to the field of biophysical virology and inspire future research in this exciting and everchanging discipline.

Acknowledgements

First, I would like to thank and acknowledge the traditional Blackfoot territory I performed my research on. I honor the Blackfoot people for their traditional ways of knowing and caring for this sacred ground, and I am grateful to all Indigenous peoples who have shaped and continue to strengthen the community I am a part of.

I would like to express the utmost gratitude all individuals who have supported and guided me throughout the journey of completing my PhD. I extend my deepest appreciation to my supervisor, Dr. Trushar Patel, for his invaluable guidance, patience, and mentorship throughout this research endeavor. Dr. Patel's expertise and drive have significantly shaped my understanding and approach to scientific inquiry, and I am profoundly grateful for his continuous support. You have been incredibly understanding and patient throughout my graduate school experience, and I cannot thank you enough for your direction, as well as your embodiment of hard work.

I am also grateful to my committee members, Wade Abbott, and Tony Russel, for their insightful feedback, constructive criticism, and invaluable contributions to the optimization of my project. Their encouragement and guidance have been pivotal in refining my research methods and enhancing the quality of my results. They have been continually supportive and encouraging me of my aspirations in RNA virus research, and their commitment to nurturing my academic and professional growth does not go unnoticed.

I would like to extend my sincere appreciation to my supportive lab group, whose camaraderie and collaborative spirit have enriched my research experience. Special thanks to Liam and Higor for their assistance in experimental design and lab work, which have been integral to the success of my project. Their expertise and dedication have been invaluable assets to my

research endeavors. Liam, you have been such an important and supportive person in my life since you have joined our lab. Thank you for always listening to me and thank you for being such a great person to work alongside. I will always cherish our friendship.

Aysha, I am beyond grateful for your remarkable resilience and support for me as I balanced my research and personal life events. Your unwavering support, friendship, and understanding have been a source of inspiration, and I am eternally thankful for your friendship and solidarity.

Jenna, I am profoundly grateful for your unwavering emotional support and friendship throughout this journey. Your kindness and encouragement have been a source of strength during challenging times, and I am truly fortunate to have you in my life.

I would like to extend my heartfelt appreciation to all the remarkable women, and all others in my lab and personal life for their unwavering support and acceptance during my journey in graduate school. Their kindness, empathy, and inclusivity have been a beacon of light, and I am immensely grateful for their friendship and support.

I would like to thank my parents for their love and acceptance toward me. Thank you for words of encouragement during my overall post-secondary experience, as well as being supportive individuals no matter what. I am forever grateful for you two. Lastly, I would like to thank Gabe for being such an incredible man in my life. Thank you for your curiosity and passion for learning about my research, being an amazing partner and teammate, and as well, that you treat me with the utmost respect and support me no matter what.

Table of contents

List of Tables

List of Figures

List of abbreviations

1-D – One-dimension

3-D – 3-dimension

AAV – Adeno-associated virus

AUC – Analytical ultracentrifugation

ATP – Adenosine triphosphate

ATSAS – Automated transmission system for small-angle X-ray scattering

BLASTn – Nucleotide basic local alignment tool

BLIP – β-lactamase TEM-1 inhibitory protein

BrUTP – Bromo-UTP

CDC – Centers for disease control and prevention

cccDNA – Covalently closed circular DNA

CD spec – Circular dichroism spectroscopy

cGcG – Consecutive G over consecutive C ratio

Comp – Compensatory

CpG – Cytosine and guanine dinucleotides

CRISPR – Clustered regularly interspaced short palindromic repeats

Cryo-EM – Cryogenic electron microscopy

CS – Cyclization sequence

DAR – Downstream of the AUG region

DCS-PK – Downstream of the cyclization sequence pseudoknot

DDX17 – DEAD-box ATP-dependent RNA helicase 17

DDX3X – DEAD-box ATP-dependent RNA helicase 3X

DENV – Dengue virus

DIG – Digoxigenin

DMP – Dimethyl pimelimidate

Dmax – Maximum particle dimension

DNA – Deoxyribonucleic acid

DRC – Democratic Republic of the Congo

dsRBD – Double-stranded RNA-binding domains

FMDV – Foot and Mouth disease virus

Fnorm – Fluorescence normalized

FXR1 – Fragile X mental retardation syndrome-related protein 1

G4 – G-quadruplex

GBS – Guillain Barré syndrome

HCV – Hepatitis C virus

HCWT – High cyclization wild type

HeLa – Henrietta Lacks

HERV-W – Human Endogenous Retrovirus-W

hnRNP – Heterogeneous Nuclear Ribonucleoprotein

HPLC – High-pressure liquid chromatography

Hz – Hertz

IGR – Intergenic region

ILF2 – Interleukin-enhancer binding factor 2

ILF3 – Interleukin-enhancer binding factor 3

IP – Immunoprecipitation

IPTG – Isopropylthio-β-galactoside

IR – Infrared radiation

IRES – Internal ribosome entry site

ITC – Isothermal titration calorimetry

IVT – In vitro transcription

JEV – Japanese encephalitis virus

Kd – Dissociation constant

KH – K-homology domains

LC-MS – Liquid chromatography – mass spectrometry

lncRNA – long non-coding RNA

M6A – N6-methyladenosine

MALS/DLS – Multi-angle light scattering/dynamic light scattering

MBF – Mosquito-borne flaviviruses

MP1 – Mpox 1

MP2 – Mpox 2

MPXV – Mpox virus

Mut – Mutant

MW – Molecular weight

MST – Microscale thermophoresis

MVEV – Murray Valley fever virus

NCL – Nucleolin

NCR – Non-coding region

NMR – Nuclear magnetic resonance

nm – Nanometres

NTP – Nucleotide triphosphate

P(r) – Pair-wise distribution

PAGE – Polyacrylamide gel electrophoresis

p53 – Transformation-related protein 53

PANTHER – Protein analysis through evolutionary relationships

PDS – Pyridostatin

PHQS – Putative hybrid-quadruplex sequence

piRNA – PIWI-interacting RNAs

PowV – Powassan virus

PSL2 – Packaging stem-loop 2

PQS – Potential G-quadruplex sequence

QGRS – Quadruplex forming G-Rich sequences

RA – Response amplitude

RBPs – RNA-binding proteins

RdRp – RNA-dependent RNA polymerase

RFP – Red fluorescent protein

Rh – Radius of hydration

Rg – Radius of gyration

RNA – Ribonucleic acid

RNPs – Ribonucleoproteins

RRM – RNA recognition motif

rRNA – Ribosomal RNA

SAXS – Small-angle X-ray scattering

SDS – Sodium dodecyl sulfate

SEC – Size exclusion chromatography

SELEX – Systematic evolution of ligands by exponential enrichment

sfRNA – Subgenomic flaviviral RNA

SHAPE – Selective 2’-hydroxyl acylation analyzed by primer extension

sHP – Short hairpin

sHP-3’SL – Upstream of the small hairpin 3’ stem-loop

SLA – Stem loop A

SL – Stem loop

SLB – Stem loop B

SnaR-RNA – Small NF90 associated RNA

SNR – Signal-to-noise ratio

ssDNA – Single-stranded DNA

ssRNA – Single-stranded RNA

STRING – Search tool for the retrieval of interacting genes/proteins

TBF – Tick-borne flaviviruses

TEV – Tobacco Etch virus

ThT – Thioflavin T

TMPyP4 – 5,10,15,20-Tetrakis-(N-methyl-4-pyridyl) porphine

TO – Thiazole orange

TRIC – Temperature related intensity change

tRNA – Transfer RNA

UAR – Upstream of the AUG region

UFS – 5’ UAR flanking system

UTP – Uracil triphosphate

UTRs – Untranslated terminal regions

UV – Ultraviolet radiation

WNV – West Nile virus

YFV – Yellow fever virus

ZAP – Zinc finger CCCH-type antiviral protein 1

ZIKV – Zika virus

Chapter 1. Introduction and background

Viruses are obligate intracellular parasites comprising a mobile genetic element that relies on the fate of a cell to generate copies of themselves [1]. They predominantly contain a metastable protein capsid that encapsulates and stabilizes the nucleic acid genome, which allows for the transport of the genome, and viral tropism, which gives rise to binding specificity and import into susceptible cells [2]. However, a viral genome inside of a cell is not guaranteed a successful replication, as a permissive environment is required for the viral genome to gain an advantage over antiviral defenses and highjack specific host proteins that are critical in amplification of the viral genome [3]. The exploitation of a suitable host cell by the genome requires two criteria - an optimal landscape of host biomolecules, and a successful genomic sequence/three-dimensional structure of the viral genome that gives rise to its pathogenesis and hijacking capabilities [4-6]. Through modern molecular biology and sequencing capabilities, viruses have been revealed to infect all domains of life, and have coevolved alongside their hosts over time, insomuch that it's been revealed that many other organisms living in the same ecosystem rely on these pathogens for survival [7]. Recent studies have shown that several identified viral species have beneficial, even critical symbiotic elements to the survival and development of higher-order organisms [8]. The process of a virus coevolving with its host that then provides a new function that the host then relies on is called symbiogenesis [9]. This is seen in a parasitoid wasp, wherein they inject their eggs into certain caterpillar species, along with polydnavirus virions [9]. These virions are critical for the survival of the wasp eggs, as the proteins expressed by the virus prevent the caterpillar from encapsulating the wasp eggs, thus destroying them [9]. Moreover, examples of this can also be seen in humans, wherein an endogenous retrovirus integrated its genome into a common ancestor of mammals [10-12]. This virus, known as Human Endogenous Retrovirus W (HERV-W), expresses a protein called Syncytin, which has become a critical protein for the development of the placenta, and the survival of a developing fetus [10-12]. In summary, viruses provide a myriad of functions, positive or negative to the host and global ecosystems, but all these intricate functions they possess for pathogenesis are not possible without the virus maintaining specific sequences and overall structure that allows for intracellular permissivity.

Research into viruses today has been mostly driven by socioeconomic factors to ensure the health of our species, or manipulation/exploitation of environmental material which in turn, drives economic stability and longevity of life [13-15]. Furthermore, many industries such as agriculture and livestock, research heavily into pathogens, as disease impacts revenue and the health of humans [16, 17]. There is a plethora of research findings yet to be discovered that explain the intricacies of the hundreds upon thousands of viral species and the impact they contribute to ecosystems [18-20]. Because scientific funding is heavily tied to political policy, economics, and decreasing human suffering, the elucidation of 451-513 virus species currently identified to infect humans as of 2025 has been ongoing, more so than the hundreds of thousands that remain to be fully discovered on the planet that each provided critical insights to viral processes [18-26]. Furthermore, the research conducted so far of these ~451-513 viruses that possess the capability of infecting humans comprises 54 families of viruses [27]. It has been shown that RNA viruses tend to be the majority of the viral species that infect humans vs. DNA viruses, and this difference is a result of RNA viruses having a greater number of mutations they acquire during replication inside a host, allowing epitope variation, thus decreasing rapid immune response [28, 29].

With virology research primarily dedicated to viruses that infect humans, many avenues are being investigated to target them and attenuate/eliminate viral replication. Some targets of interest include the investigation of viral genomic structures as well as the host-protein machinery that is exploited by these structures during replication [30-37].

1.1 Flavivirus threat to public health

Within the range viral families that infect humans, the taxonomic family of viruses known as *Flaviviridae* infect a wide range of vertebrates and invertebrates terrestrially and in marine ecosystems [38-43]. This family can be subdivided into 4 genera; *Hepacivirus, Pegivirus,* and *Pestivirus,* which so far have only been identified in mammalian hosts, and lastly, the genus *Flavivirus* [42, 43]. The *Flavivirus* genus is unique in that it infects a wide range of vertebrates and invertebrates terrestrially and in marine ecosystems [39-43]. Notably, a collection of flaviviruses known to cause disease in humans have become public health concerns, and more focus has been devoted to understanding their structure and pathogenicity than flaviviruses that infect other organisms [38, 44, 45]. These flaviviruses are known as mosquito-borne flaviviruses (MBFs) and tick-borne flaviviruses (TBFs), and they possess the ability to circulate between their primary arthropod vectors and vertebrates, albeit under stringent control of mutations in specific structures critical for efficient host-switching and infection in both hosts [45-47]. MBFs and TBFs have a significant public health impact across the globe causing mild fever to severe neurological complications and death. In 2021, there were 59 million cases of Dengue virus (DENV), with outbreaks continuing in over 110 countries, 86,500 yellow fever cases, and 170,000 Zika virus (ZIKV) cases [48, 49]. While most infections are mild or asymptomatic, severe cases can result in hospitalization, long-term disability, or death [48, 49]. The rapid interplay of socioeconomic globalization, urbanization, and climate change has created unprecedented opportunities for MBFs/TBFs to expand into new regions [48-56]. Expanding on these examples, with the increase and reliance on international travel and trade, such as the inadvertent transport of *Aedes* mosquitoes or *Ixodes* ticks via cargo shipments, enabling these vectors to establish footholds in previously unaffected areas [50-56]. Urbanization further causes the expansion of these pathogens by concentrating human populations in dense cities, as well as agricultural engineering of land causing migration of these organisms into urban centers [50-56]. Lastly, climate change has allowed the expansion of the geographical range of where *Aedes* mosquitoes or *Ixodes* ticks can reside and harbour MBFs & TBFs, respectively [50-57]. For example, rising temperatures in North America have been linked to the northward expansion of *Ixodes scapularis* ticks, driving increased human exposure to Powassan virus – a TBF once confined to limited regions but now reported by the CDC and Health Canada to be as far north as Canada [50-56]. Similarly, erratic weather patterns create ideal breeding conditions for mosquitoes, amplifying arboviral outbreaks. Together, these interconnected drivers threaten biodiversity by disrupting native species' living patterns and allows spillover events into new vertebrate hosts [50-52]. Akin to other MBF's, ZIKV, a virus prioritized by the World Health Organization for research and development, causes a host of neurological sequelae in adults, is sexually transmitted up to 9 months after infection, and can cause gestational microcephaly from vertical transmission of an infected pregnant individual to their developing fetus [58-60]. Moreover, the two main species susceptible to carrying ZIKV and circulating it to vertebrates are *Aedes aegypti* and *Aedes albopictus* [58-60].

The current threat of Poxvirus outbreaks in humans

The DNA virus family, Poxviruses, present a substantial threat to the planet, especially with the resurgence of Mpox coupled with vaccine hesitancy from misinformation [61-65]. Mpox, like Smallpox, is caused by a virus in the Orthopoxvirus genus, but while Smallpox had undetectable cases in the 1980's due to vaccination efforts, Mpox continues to circulate among humans and animals [66-68]. Recently, Mpox has become a global concern due to its ability to spread across continents because of the rapid movement of humans and goods with trade and air travel, causing over 100,000 cases in more than 120 countries, including regions where it was previously unknown [66-68]. The World Health Organization has classified the current Mpox outbreak as a public health emergency of international concern, given its rapid expansion and emergence of new clades, notably in Africa and beyond [66-68].

Comparing and contrasting both Mpox and Smallpox, both can cause fever, rashes, and severe complications, but the fatality rate of Mpox (1–10%, depending on the clade) is much lower than that of Smallpox (30–50%) [66, 68, 69]. Unlike Smallpox, Mpox can infect a broader range of animal hosts, making containment more challenging [66, 68, 69]. Smallpox vaccination provides some protection against Mpox, but with waning immunity due to deceased Smallpox vaccination that produced high global immunity Mpox cases have risen [66, 67, 70]. Although current Mpox outbreaks are generally less lethal, the ongoing evolution of the virus and its zoonotic potential pose a continuous threat that stresses the need for continuous surveillance, vaccination outreach, and international public health coordination – especially as global interconnectedness accelerates the risk of widespread outbreaks [61-65, 67, 69, 70].

ZIKV and Mpox virus life cycles

ZIKV is a positive-sense RNA genome that is ~11 kilobases, with 3 structural genes and 7 non-structural genes, conjointly flanked by long, conserved 5' and 3' untranslated terminal regions (UTRs) [71-73]. Infection begins when an infected female *Aedes* mosquito initiates a blood feed on a host containing the optimal biochemical landscape [71-73]. The viral envelope (E) protein mediates attachment to receptors on host cells, such as TAM receptors AXL, Tyro3, TIM1, and Dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin (DC-SIGN), facilitating viral entry via clathrin-mediated endocytosis [71-73]. After uptake, the viral capsid performs a pH-dependent conformational change and releases the single-stranded RNA genome into the cytoplasm [71]. The positive-sense RNA serves as the template for the ribosome and is translated attached ribosomes to the endoplasmic reticulum (ER) into a single polyprotein [71]. This polyprotein is cleaved into structural proteins (capsid, membrane, and envelope) needed for virion assembly, and non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5) critical for RNA replication and modulation of the entire cellular landscape [71]. Replication occurs on rearranged endoplasmic reticulum (ER) membranes the NS5 polymerase synthesizes a negative strand intermediate, which serves as a template for new positive strands [71]. Newly formed positive sense genomes are packaged and coordinated into the organized structural proteins which altogether bud into the ER lumen [71]. Virions are then transported through the ER-Golgi intermediate compartment and the Golgi apparatus, where glycosylation/maturation occurs, before its extracellular release via exocytosis [71].

The tissues that ZIKV has been shown to predominantly infect are neurological tissues, including neural progenitor cells, astrocytes, and microglia, leading to neurodevelopmental defects like microcephaly due to disruption of cell cycle, induction of apoptosis, and neuroinflammation [74]. The virus also modulates host immune pathways, including inhibition of apoptosis, allowing life extension of the cellular environment for the virus, and consequently, more efficient propagation to neighbouring host cells [74].

Mpox Life Cycle and Spread

Mpox is a double-stranded DNA virus belonging to the Orthopoxvirus genus in the Poxviridae family. The genome is approximately 197,000 base pairs in length, and so far, has been shown to contain 181 protein coded genes [75, 76]. Mpox enters host cells through fusion of the viral envelope with the plasma membrane or by endocytosis and macropinocytosis pathways [75, 76]. Once inside, the viral genome is released into the cytoplasm wherein transcription of early genes occurs using virus-encoded RNA polymerases, independent of the host nucleus [75, 76]. The early gene products made modulate host defenses and prepare the cell for viral DNA replication, which takes place in cytoplasmic viral replication factories [75, 76]. Subsequent intermediate and late genes are then transcribed, producing structural proteins and enzymes needed for virion assembly [75, 76]. The viral DNA is then replicated by its own DNA polymerase [75, 76]. Immature virus particles assemble in ‘cytoplasmic factories’ and mature into functional virions [75, 76].

Regarding the spread of Mpox, it traditionally transmits from animals (primarily rodents and primates) to humans via direct contact or bites, but regarding human-to-human transmission, it transmits through respiratory droplets, contact with lesions, or contaminated material [77, 78]. Recent outbreaks showed an increased transmission due to multiple variables, such as close contact in social encounters, urbanization, and waning Smallpox immunity post-vaccination era [77, 78].

Together, the ZIKV virus and Mpox virus exemplify different viral replication strategies aligned with their specific genomic landscapes and diverse modes of transmission influenced by ecological and human behavioral factors.

Current vaccination efforts for ZIKV and Mpox viruses

Vaccine development with ZIKV has been a challenge due to fears of it triggering Guillain-Barré syndrome, which is an autoimmune condition in which an individual's immune system mistakenly attacks the peripheral nerves, leading to nerve damage that affects muscle movement as well as sensations such as pain, temperature, and touch [79-83]. Moreover, it also poses issues of antibody-dependent enhancement (ADE) of infection with DENV – This phenomenon is a process where antibodies from a prior infection or vaccination bind to a virus (DENV in this case) without neutralizing it, instead helping DENV to enter and infect host cells more efficiently [79-86]. This can lead to increased viral replication and more severe disease, making ADE a critical concern in vaccine development [79-86]. Several therapeutics targeting ZIKV have progressed from preclinical development into clinical trials. Regarding vaccines, Moderna's mRNA-1893 ZIKV vaccine has completed Phase 1 trials and moved into Phase 2, demonstrating robust immunity in participants [87]. Takeda's purified inactivated vaccine (TAK-426) and the NIAID's Zika DNA vaccine (VRC5283) have also shown positive results in Phase 1 trials, with TAK-426 advancing further and VRC5283 demonstrating strong immunogenicity [87]. In addition to vaccines, small molecule inhibitors such as tetracyclines (e.g., methacycline, minocycline) and MK-591, as well as other compounds identified through in silico studies (NITD, compound-6, P02, Doxytetracycline, Rolitetracycline), have shown potent antiviral effects in preclinical models [88, 89]. Currently there are no ZIKV therapeutics approved, but the vaccine candidates and lead compounds reflect ongoing efforts to combat this virus.

Conversely with Mpox virus, the vaccine currently in use is the Imvamune® (also known as Jynneos® in the US, Imvanex® in Europe, or MVA-BN globally), which was primarily designed for the Variola virus (Smallpox) [75, 77, 78, 90-92]. There is crossover immunity with the Imvamune vaccine for Mpox because it is close genetic relationship with Smallpox, and the vaccine is protective against both Smallpox and Mpox [75, 77, 78, 91, 92]. Studies have shown that using the Smallpox vaccine can be about 85% effective in preventing Mpox infection [75, 77, 78, 91, 92]. The vaccine triggers both antibody and cellular immune responses that confer protection, although the exact mechanisms of protection against Mpox are still being studied [75, 77, 78, 91, 92].

Host-viral interaction studies

Narrowing on the topic of host-viral interfaces that are critical for a virus's propagation, there are many well-characterized examples of host protein-viral RNA/DNA interactions. These interfaces contain critical sights that shape the infection strategies that these viruses utilize for replication, immune evasion, and host-switching. Several recent studies provide direct biophysical characterization and kinetic measurements of these interactions. Particularly, McKenna et al. (2006) presented a comprehensive biophysical analysis of the interaction between human protein kinase R (PKR) and viral inhibitor RNAs from Epstein-Barr virus and adenovirus [93]. They measured binding affinities (K_d) of using equilibrium fluorescence anisotropy, which measures interactions in solution by measuring rotational freedom of a fluorescently labeled molecule, specifically showing that human PKR binds to different RNAs, including viral inhibitor RNAs derived from Epstein-Barr virus (EBER-1 and EBER-2), adenovirus VAI RNA, as well as synthetic double-stranded RNA. They further measured the rapid kinetic association rates using stopped-flow fluorescence and quantified how binding kinetics relate to functional PKR activation [93]. Their findings revealed distinct kinetic properties for inhibitor RNAs versus standard dsRNA, tightly linking RNA-protein binding parameters to inhibition of antiviral signaling [93]. Another example of characterization of host-viral interactions can be seen with Schmidt et al. (2021) and Flynn et al. (2021) [94, 95]. Both groups elucidated the landscape of host–viral RNA interactions during SARS-CoV-2 infection by combining RNA antisense purification with mass spectrometry (RAP-MS or the analogous ChIRP-MS). These methods allowed the identification of proteins that directly associate with viral RNAs in infected cells. The authors utilized UV crosslinking to stabilize RNA–protein complexes prior to purification. The interactions were further validated using biochemical assays such as electrophoretic mobility shift assays (EMSAs), competitive binding experiments, and additional crosslinking approaches, collectively revealing a diverse network of host proteins with potential roles in viral replication, immune modulation, and antiviral defense

[94, 95]. These studies provided key information for understanding the mechanistic basis of SARS-CoV-2 host interactions, offering critical insights for future therapeutic strategies [94, 95].

An example of a DNA virus interaction with human proteins involves the human host DNA sensors interferon-γ inducible protein 16 (IFI16) and cyclic GMP-AMP synthase (cGAS) in detecting and responding to viral DNA from poxviruses and herpesviruses [96-98]. IFI16 and cGAS are key innate immune proteins that directly bind viral double-stranded DNA (dsDNA) in a sequence-independent manner, initiating antiviral signaling pathways [97]. These research groups used surface plasmon resonance (SPR), a label-free technique that measures refractive index changes to a ligand binding to a immobilized substrate, as well as fluorescence polarization (FP), which measures changes in the rotational mobility of a fluorescently labeled molecule upon binding to a larger partner, detecting molecular interactions by quantifying the resulting increase in emitted light polarization [96, 98].

SPR experiments were used to measure the binding affinities of IFI16 to double-stranded viral DNA, showing tight binding in the sub-nanomolar to low-nanomolar range [98]. The FP assays were used to assess IFI16 binding to DNA oligonucleotides of varied lengths and structures [96].

Biophysical characterization of viruses is critical, as it provides precise, quantitative insights into the strength, kinetics, and mechanisms of molecular interactions – information essential for understanding biological function, validating targets, and guiding the rational development of therapeutics with optimized efficacy and safety. In my research, I primarily focused efforts into ZIKV as well as Mpox viruses, highlighting biophysical characterization in both RNA and DNA viruses.

ZIKV 5' and 3' untranslated terminal regions

MBF and TBFs infect a wide range of hosts and cell types, altogether replicating within the confines of the cytoplasm of the cell [99]. The adaptability and mechanisms that MBFs and TBFs utilize in host-switching are embedded within their RNA genomes, notably, within the UTRs [46, 100]. All MBFs and TBFs possess these UTRs, and the ZIKV UTRs are known to harbour many sequences and structures that play critical roles in balancing efficient viral transcription and translation, host adaptation, and immune evasion [46, 100-106]. These highly structured and conserved regions are involved in commandeering host cellular proteins that further assist in the propagation of the virus directly and allow the overall kinetic landscape of the cell to change and favour viral replication (Fig. 1.1) [107, 108].

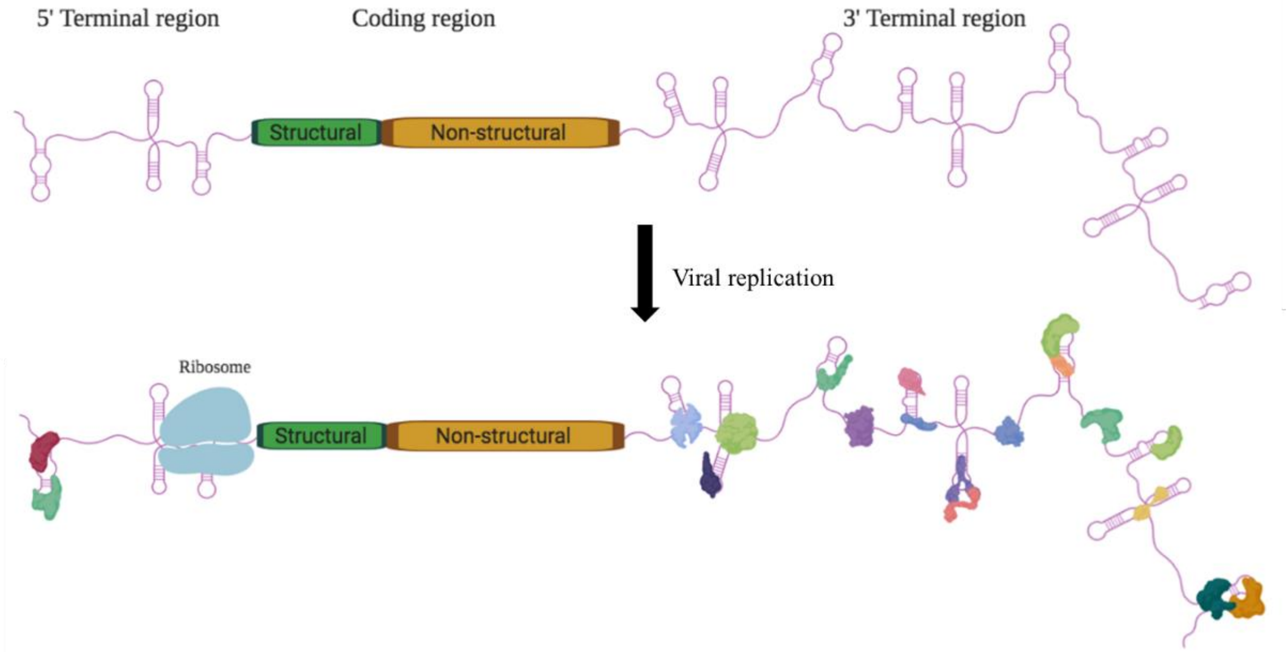


Figure 1.1. A simplistic representation of the ZIKV UTRs during viral replication. During this process, host proteins are hijacked, and the proposed functions of these events are to aid in viral replication, sequester host pathways, or prevent immune response signalling in the cell (created with BioRender.com).

The human proteins that are hijacked by the UTRs and used to leverage viral replication in ZIKV have not been fully identified and have yet to be elucidated on where and how they bind to the UTRs, whether directly or indirectly [109-112]. I aimed to identify these binding partners to provide insight into the identification of the human proteins being hijacked but also provide an avenue for novel therapeutics.

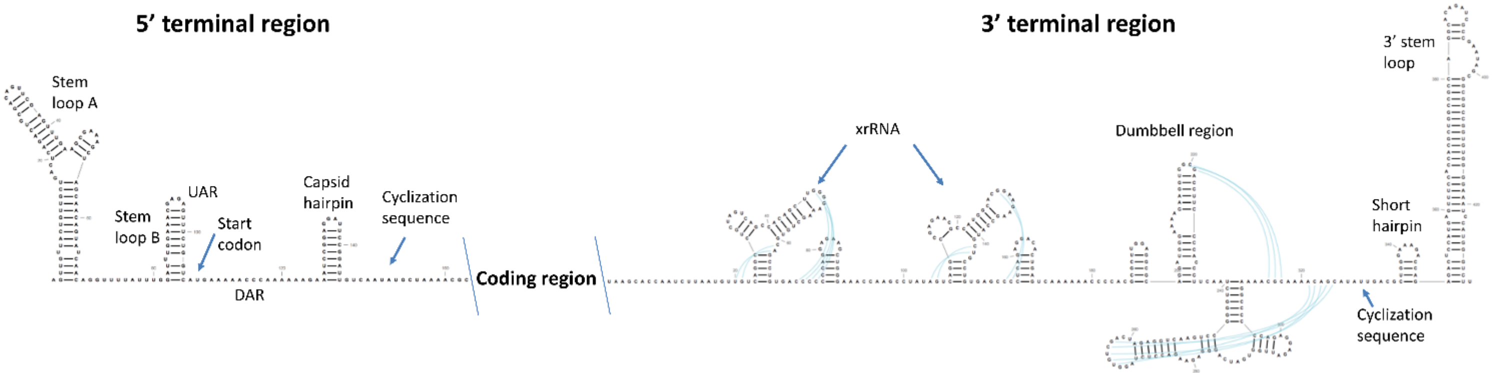


Figure 1.2 The secondary structure topology of the ZIKV UTRs that were collectively determined by SHAPE [113], along with the sequences involved in cyclization of the viral genome to regulate transcription and translation of its genome. After the start codon, the 5' cyclization sequence is located in the capsid gene of the virus which is indicated. The remaining coding sequence for all other genes are contained in the 'coding region' section.

Human Nucleolin (NCL) identified as a binding partner with ZIKV terminal regions

In my research findings, NCL has been identified as a key cellular factor that binds to the ZIKV TRs (see Ch. 4/5). The relationship between NCL with viral RNA is well documented, and so far, shows evidence of aiding in viral replication and assembly of virus particles [114, 115]. NCL is a multifunctional phosphoprotein abundantly expressed in human cells, and is crucially involved in ribosome biogenesis, chromatin remodeling, and RNA metabolism [116]. Structurally, NCL is composed of three principal regions: An N-terminal acidic domain enriched in phosphorylation sites for protein-protein and chromatin interactions [117], a central region containing four tandem RNA Recognition Motifs (RRMs) which mediate high-affinity and sequence-specific RNA binding [117], and finally, a C-terminal end that contains a glycine/arginine-rich domain (GAR or RGG domain) that contributes to RNA and protein interactions, specifically, G4 binding, and subcellular localization [118]. Focusing on the RRMS, each domain consists of about 80 – 90 amino acids forming four anti-parallel beta strands packed against two alpha helices (βαββαβ sandwich) that creates a stable platform for binding RNA with sequence specificity. The RRMs are highly conserved across prokaryotes and eukaryotes, showing evolutionary pressure to maintain RNA-binding function due to its multifunctional biological importance [118]. The conserved βαββαβ sandwich folding creates a surface optimized for binding single-stranded RNA [117]. These RRMs engage RNA primarily through stacking interactions of conserved aromatic residues with nucleobases and electrostatic interactions with the RNA backbone [118]. In NCL, the RRMs cooperate to achieve versatile RNA binding specificity and affinity, capable of recognizing diverse RNA sequences and structures with a strong preference for guanine-rich sequences, enabling it to specifically bind and modulate G-quadruplex (G4) structures [118]. Briefly, G4s are nucleic acid secondary structures formed within DNA and RNA [119-121]. They are characterized by the stacking of planar guanine tetrads, in which four guanine bases are held together by Hoogsteen hydrogen bonding [119-121]. The overall structure can vary in loop size and topology, but the Hoogsteen base pairing is specific to the guanine residues within each tetrad (Fig. 1.3) [119-121]. High-resolution structural research on NCL, showed that it binds to the human DNA G4 in the MYC oncogene promoter, reveal that NCL's RRMs engage G4s through multivalent interactions involving loops and flanking regions of the G4, stabilizing these structures and influencing gene expression regulation [122]. In the context of viral infections, both Mpox and ZIKV genomes harbor G4 structures (see Ch. 6 & 7). NCL and its RRMs form a connection between host RNA-binding proteins and viral nucleic acid secondary structures, particularly G-quadruplexes, and understanding the biophysical characteristics of NCL-G4 interactions in viral contexts provides insight into virus-host interplay and opens avenues for antiviral strategies targeting these structured nucleic acid motifs.

1.4 G-quadruplex structures as an exploitative target for antiviral therapies in ZIKV and Mpox viruses

In the past, nucleic acids have been shown to predominantly fold and stabilize via traditional Watson-Crick base pairing [123]. Since Watson and Crick published their article in 1953, further elucidation of non-traditional, known as non-canonical structures in both DNA and RNA has been discovered, which include: Z-DNA, A-DNA, G-quadruplexes (G4), triplexes, cruciform, i-motifs, hairpins in DNA and RNA, and pseudoknots in RNA [124]. Furthermore, these structures were found to be in every organism and virus that have been investigated for non-canonical structure [125, 126]. I specifically investigated viral G4 structures, as well as host-proteins being hijacked by them to benefit viral replication. G4s have become a prospective target of interest for us not only for ZIKV, but also for Mpox (formally known as Monkeypox), as both need immediate alternative antiviral targeting for countries without vaccine infrastructure.

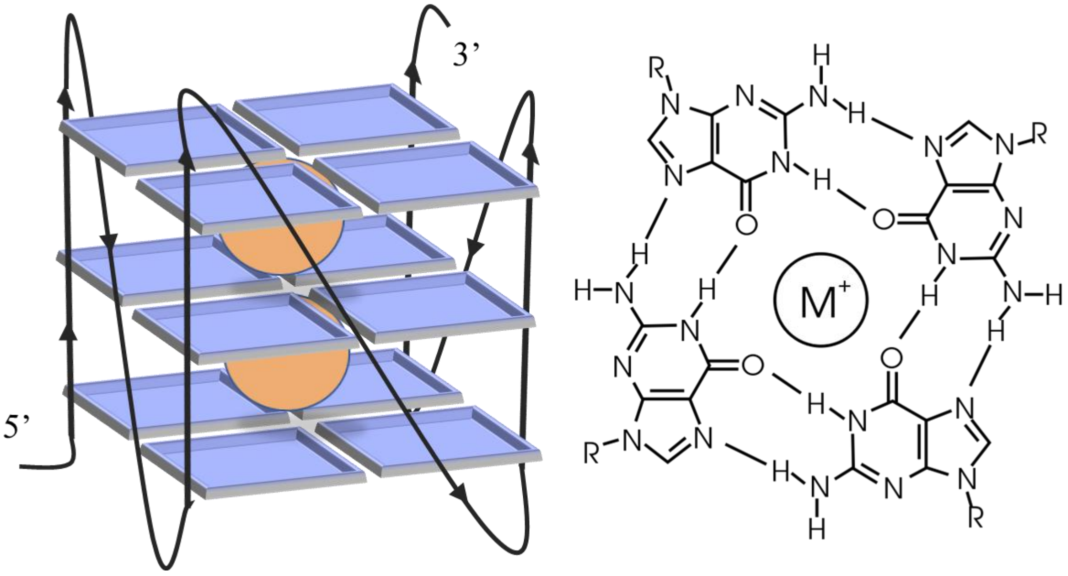


Figure 1.3 A G4 schematic showing the guanine stacking (left), as well as a the top/bottom view presenting the Hoogsteen base pairing of the guanines with a monovalent cation coordinated for stabilization of the non-canonical structure (right).

In the context of viral G4s, Hepatitis C virus (HCV) was shown to possess a G4 on its internal ribosome entry site (IRES) within its 5’ UTR [127-132]. The research group that discovered this suggested that it could be used to suppress viral translation [127, 128]. Furthermore, another G4 contained in the core gene of HCV showed that when NCL binds and stabilizes the G4, it prevents translation from occurring [127-129]. Another example is seen in the Hepatitis B virus (HBV), where it was discovered that the G4 in the pre-core promotor of covalently closed circular DNA (cccDNA) is critical for the transcription of the C-gene, and consequently, translation of the viral core proteins [133]. These examples show that G4s could possess multi-functional roles depending on the virus and region it is located within the viral genome. As it stands to date, there are ~2800 G4 binding molecules, with many in clinical trials [86]. Previous published data shows the need for developing novel antiviral therapies to target viruses wherein there is no vaccine, or areas with poor vaccine infrastructure, with G4s being one of those potential targets [134, 135].

With this approach, researchers can either block host protein-viral G4 interactions, or unfold or stabilize the viral G4, preventing functionality, and hopefully attenuating the infection of the virus [37, 134].

This thesis identifies human RNA-binding proteins that interact with ZIKV genomes, specifically, the terminal regions. From this data I biophysically characterize hypothesized critical interactions with ZIKV that we investigate that could provide insight into antiviral targeting. Furthermore, I also investigated the relationship of ZIKV and Mpox G4s with human proteins in hopes to allow for an alternative target for viral attenuation. Chapters 2 discusses the core techniques that were used to measure and characterize the protein-nucleic acid interactions, and as well, discusses the method of microscale thermophoresis, which allows for the elucidation of binding between two species, as well as determining a dissociation constant (K_d) of the interaction. This was performed on the ZIKV 5' and 3' UTRs, as well as their interactions with human proteins (Chapters 4, 5, and 6). Furthermore, this was also performed on studies with the Mpox virus interacting with a human protein, as well as a potential antiviral treatment option for Mpox infection (Chapter 8). Chapter 2 also discusses the technique of small-angle X-ray scattering (SAXS), which gives insights into the overall solution structure of our biomolecules under study. Specifically, I utilized SAXS for elucidating the structure of the 5' and 3' UTRs of ZIKV, and the human protein NCL, altogether to biophysically characterize the interactions between them (Chapter 5). Furthermore, I also utilized SAXS to understand if a G4 was present within the Mpox genome, and folds in solution (Chapter 8). Chapter 2 discusses the techniques I utilized in my experiments, including circular dichroism spectroscopy, which allows for the detection of secondary structure formation of biomolecules (for our studies) based on the absorbance of ultraviolet circular polarized light. I utilized this technique as a complementary method to SAXS to identify if the Mpox G-quadruplex was present. At also discusses multi-angle/dynamic light scattering which gave us insight into the solution biophysics of the various biomolecules under study. These include understanding the solution behaviour of NCL to study its interactions with the ZIKV virus, as well as the UTRs of ZIKV to study the interactions that they make with each other (see Chapter 5). Lastly, for methods, Chapter 4 discusses the capture and identification of proteins, directly and indirectly, interacting with bait RNA using various techniques, including small molecule modifications on the RNA, aptamers, nucleotide substitution, or DNA anchoring. This is a review article that is published in the Canadian Science Publishing - Journal of Biochemistry and Cell Biology.

Chapters 4-8 focus on the experimental findings, starting with Chapter 4 – the identification of RNA-binding protein partners interacting with the 5' and 3' TRs of the ZIKV virus. Using a digoxigenin small molecule immunoprecipitation assay approach I used the 5' and 3' UTRs as bait RNA, as well as the neuro-like cell line SH-SY5Y to give the results neuroptropistic relevance due to ZIKV virus' neuroinvasive replication. I identified several binding partners that I used as our template for investigating and elucidating the replication dynamics of ZIKV virus, and equally important, laid a foundation for prospective targets using novel antiviral therapies. Chapter 5 is the study of one of the identified proteins from the pulldown assay, NCL, and its characterization of the interactions with the ZIKV 5' and 3' UTRs. In this chapter, I verify that NCL does bind directly to the UTRs of the virus with high affinity in the sub-nanomolar range, has binding affinity variation between the 5' and 3' UTRs, as well as variation within regions of the 3' UTR itself. Moreover, I elucidate a novel discovery of NCL's criticality for ZIKV replication and show using a knockdown approach that NCL is crucial for the formation of infectious virus particles as well as generating viral RNA transcripts. Chapter 6 is an accepted article in the Canadian Science Publishing - Journal of Biochemistry and Cell Biology, which is a study of a human protein called DDX17 that was identified in the ZIKV UTR pulldown assays, which specifically interacts with the 3’ UTR and unfolds the G4 contained in the 3’ SL. Chapter 7 focuses on the cyclization interactions that the ZIKV UTRs make with themselves. I elucidated as well that ZIKV has variation in cyclization kinetics *in vitro* between the Asian and African strains. Finally, Chapter 8 concludes with a publication in the Journal of Medical Virology on the identification and characterization of two G4 contained in the Mpox genome. Using biophysical methods, I show that these regions fold into quadruplexes, and that treating infected human cells with the G4 binding small molecule, TMPYP4, significantly stifled viral protein production, showing a potential avenue for novel antiviral therapeutics.

1.5 References

1. Gelderblom, H.R., *Structure and Classification of Viruses*, in *Medical Microbiology*, S. Baron, Editor. 1996, University of Texas Medical Branch at Galveston
Copyright © 1996, The University of Texas Medical Branch at Galveston.: Galveston (TX).
2. Mateu, M.G., *Assembly, stability and dynamics of virus capsids.* Arch Biochem Biophys, 2013. **531**(1-2): p. 65-79.
3. Segondy, M., *[Host specificity of viruses and inter-species crossing].* Rev Francoph Lab, 2010. **2010**(423): p. 37-42.
4. Lycke, E. and E. Norrby, *11 - Virus-induced changes of cell structures and functions*, in *Textbook of Medical Virology*, E. Lycke and E. Norrby, Editors. 1983, Butterworth-Heinemann. p. 93-104.
5. Fairman, C.W., A.M.L. Lever, and J.C. Kenyon, *Evaluating RNA Structural Flexibility: Viruses Lead the Way.* Viruses, 2021. **13**(11).
6. Boerneke, M.A., J.E. Ehrhardt, and K.M. Weeks, *Physical and Functional Analysis of Viral RNA Genomes by SHAPE.* Annu Rev Virol, 2019. **6**(1): p. 93-117.
7. Weitz, J.S. and S.W. Wilhelm, *Ocean viruses and their effects on microbial communities and biogeochemical cycles.* F1000 Biol Rep, 2012. **4**: p. 17.
8. Roossinck, M.J. and E.R. Bazán, *Symbiosis: Viruses as Intimate Partners.* Annual Review of Virology, 2017. **4**(1): p. 123-139.
9. Edson, K.M., et al., *Virus in a parasitoid wasp: suppression of the cellular immune response in the parasitoid's host.* Science, 1981. **211**(4482): p. 582-3.
10. Cornelis, G., et al., *Retroviral envelope syncytin capture in an ancestrally diverged mammalian clade for placentation in the primitive Afrotherian tenrecs.* Proc Natl Acad Sci U S A, 2014. **111**(41): p. E4332-41.
11. Durnaoglu, S., S.K. Lee, and J. Ahnn, *Syncytin, envelope protein of human endogenous retrovirus (HERV): no longer 'fossil' in human genome.* Anim Cells Syst (Seoul), 2021. **25**(6): p. 358-368.
12. Roossinck, M.J., *The good viruses: viral mutualistic symbioses.* Nat Rev Microbiol, 2011. **9**(2): p. 99-108.
13. Sankaran, N. and R.A. Weiss, *Viruses: Impact on Science and Society*, in *Encyclopedia of Virology (Fourth Edition)*, D.H. Bamford and M. Zuckerman, Editors. 2021, Academic Press: Oxford. p. 671-680.
14. Falagas, M.E., A.I. Karavasiou, and I.A. Bliziotis, *Estimates of global research productivity in virology.* J Med Virol, 2005. **76**(2): p. 229-3.
15. Lee, D., J. Kang, and K. Kim *Global Collaboration Research Strategies for Sustainability in the Post COVID-19 Era: Analyzing Virology-Related National-Funded Projects*. Sustainability, 2020. **12**, DOI: 10.3390/su12166561.
16. Richter, V., et al., *A systematic worldwide review of the direct monetary losses in cattle due to bovine viral diarrhoea virus infection.* The Veterinary Journal, 2017. **220**: p. 80-87.
17. Tildesley, M.J., et al., *The Role of Movement Restrictions in Limiting the Economic Impact of Livestock Infections.* Nat Sustain, 2019. **2**(9): p. 834-840.
18. Li, R., et al., *Metagenomic analysis reveals unexplored diversity of archaeal virome in the human gut.* Nature Communications, 2022. **13**(1): p. 7978.
19. Martínez Martínez, J., F. Martinez-Hernandez, and M. Martinez-Garcia, *Single-virus genomics and beyond.* Nature Reviews Microbiology, 2020. **18**(12): p. 705-716.
20. Dion, M.B., F. Oechslin, and S. Moineau, *Phage diversity, genomics and phylogeny.* Nature Reviews Microbiology, 2020. **18**(3): p. 125-138.
21. Tham, W.Y., *Science, interrupted: Funding delays reduce research activity but having more grants helps.* PLOS ONE, 2023. **18**(4): p. e0280576.

22. Laird, F.N., *Sticky Policies, Dysfunctional Systems: Path Dependency and the Problems of Government Funding for Science in the United States.* Minerva, 2020. **58**(4): p. 513-533.
23. Meirmans, S., et al., *Science policies: How should science funding be allocated? An evolutionary biologists' perspective.* J Evol Biol, 2019. **32**(8): p. 754-768.
24. Legg, T., J. Hatchard, and A.B. Gilmore, *The Science for Profit Model—How and why corporations influence science and the use of science in policy and practice.* PLOS ONE, 2021. **16**(6): p. e0253272.
25. Munari, F. and L. Toschi, *The impact of public funding on science valorisation: an analysis of the ERC Proof-of-Concept Programme.* Research Policy, 2021. **50**(6): p. 104211.
26. He, M., C.-Q. He, and N.-Z. Ding, *Human viruses: An ever-increasing list.* Virology, 2025. **604**: p. 110445.
27. Ye, S., et al., *An atlas of human viruses provides new insights into diversity and tissue tropism of human viruses.* Bioinformatics, 2022. **38**(11): p. 3087-3093.
28. Peck Kayla, M. and S. Lauring Adam, *Complexities of Viral Mutation Rates.* Journal of Virology, 2018. **92**(14): p. 10.1128/jvi.01031-17.
29. Kikkert, M., *Innate Immune Evasion by Human Respiratory RNA Viruses.* Journal of Innate Immunity, 2019. **12**(1): p. 4-20.
30. Bezzi, G., et al. *CNBP Binds and Unfolds In Vitro G-Quadruplexes Formed in the SARS-CoV-2 Positive and Negative Genome Strands*. International Journal of Molecular Sciences, 2021. **22**, DOI: 10.3390/ijms22052614.
31. Chen, N., et al., *Virus-host interaction networks as new antiviral drug targets for IAV and SARS-CoV-2.* Emerg Microbes Infect, 2022. **11**(1): p. 1371-1389.
32. de Chassey, B., et al., *Virus-host interactomics: new insights and opportunities for antiviral drug discovery.* Genome Medicine, 2014. **6**(11): p. 115.
33. Hegde, S., et al., *Inhibition of SARS-CoV-2 by Targeting Conserved Viral RNA Structures and Sequences.* Frontiers in Chemistry, 2021. **9**.
34. Kumar, N., et al., *Host-Directed Antiviral Therapy.* Clinical Microbiology Reviews, 2020. **33**(3): p. 10.1128/cmr.00168-19.
35. Leyssen, P., E. De Clercq, and J. Neyts, *Molecular strategies to inhibit the replication of RNA viruses.* Antiviral Res, 2008. **78**(1): p. 9-25.
36. Mahajan, S., et al., *Antiviral strategies targeting host factors and mechanisms obliging +ssRNA viral pathogens.* Bioorg Med Chem, 2021. **46**: p. 116356.
37. Ruggiero, E. and S.N. Richter, *Targeting G-quadruplexes to achieve antiviral activity.* Bioorg Med Chem Lett, 2023. **79**: p. 129085.
38. Cook, S., et al., *Molecular evolution of the insect-specific flaviviruses.* J Gen Virol, 2012. **93**(Pt 2): p. 223-234.
39. Porter, A.F., et al., *Novel hepaci- and pegi-like viruses in native Australian wildlife and non-human primates.* Virus Evolution, 2020. **6**(2): p. veaa064.
40. Parry, R. and S. Asgari, *Discovery of Novel Crustacean and Cephalopod Flaviviruses: Insights into the Evolution and Circulation of Flaviviruses between Marine Invertebrate and Vertebrate Hosts.* J Virol, 2019. **93**(14).
41. Lensink, M.J., Y. Li, and S. Lequime, *Aquatic Flaviviruses.* J Virol, 2022. **96**(17): p. e0043922.
42. Porter, A.F., et al., *Novel hepaci- and pegi-like viruses in native Australian wildlife and non-human primates.* Virus Evol, 2020. **6**(2): p. veaa064.
43. Bartenschlager, R. and K. Tabata, *Hepatitis C Virus (Flaviviridae)*, in *Encyclopedia of Virology (Fourth Edition)*, D.H. Bamford and M. Zuckerman, Editors. 2021, Academic Press: Oxford. p. 386-396.

44. Huhtamo, E., et al., *Novel flaviviruses from mosquitoes: mosquito-specific evolutionary lineages within the phylogenetic group of mosquito-borne flaviviruses.* Virology, 2014. **464-465**: p. 320-329.
45. Zhang, Y., et al., *Replication is the key barrier during the dual-host adaptation of mosquito-borne flaviviruses.* Proceedings of the National Academy of Sciences, 2022. **119**(12): p. e2110491119.
46. Liu, Z.-Y., et al., *Viral RNA switch mediates the dynamic control of flavivirus replicase recruitment by genome cyclization.* eLife, 2016. **5**: p. e17636.
47. Villordo, S.M., et al., *Dengue virus RNA structure specialization facilitates host adaptation.* PLoS Pathog, 2015. **11**(1): p. e1004604.
48. Qian, X. and Z. Qi, *Mosquito-Borne Flaviviruses and Current Therapeutic Advances.* Viruses, 2022. **14**(6).
49. Robertson, S.J., et al., *Tick-borne flaviviruses: dissecting host immune responses and virus countermeasures.* Immunol Res, 2009. **43**(1-3): p. 172-86.
50. Huang, Y.J., et al., *Flavivirus-mosquito interactions.* Viruses, 2014. **6**(11): p. 4703-30.
51. Diani, E., et al., *Vector-Transmitted Flaviviruses: An Antiviral Molecules Overview.* Microorganisms, 2023. **11**(10).
52. Kaaijk, P. and W. Luytjes, *Are we prepared for emerging flaviviruses in Europe? Challenges for vaccination.* Hum Vaccin Immunother, 2018. **14**(2): p. 337-344.
53. *Flaviviruses.* Perspect Med Virol, 2005. **11**: p. 13-51.
54. Pierson, T.C. and M.S. Diamond, *The continued threat of emerging flaviviruses.* Nat Microbiol, 2020. **5**(6): p. 796-812.
55. Liang, Y. and X. Dai, *The global incidence and trends of three common flavivirus infections (Dengue, yellow fever, and Zika) from 2011 to 2021.* Front Microbiol, 2024. **15**: p. 1458166.
56. Tatem, A.J., D.J. Rogers, and S.I. Hay, *Global transport networks and infectious disease spread.* Adv Parasitol, 2006. **62**: p. 293-343.
57. Organization, W.H. *Zika virus disease.* [cited 2023.
58. Carod-Artal, F.J., *Neurological complications of Zika virus infection.* Expert Rev Anti Infect Ther, 2018. **16**(5): p. 399-410.
59. Epelboin, S., et al., *Zika virus and reproduction: facts, questions and current management.* Human Reproduction Update, 2017. **23**(6): p. 629-645.
60. Mlakar, J., et al., *Zika Virus Associated with Microcephaly.* New England Journal of Medicine, 2016. **374**(10): p. 951-958.
61. Suleiman, A.K., *Knowledge, attitudes, and practices toward Mpox and vaccination: a cross-sectional study in Saudi Arabia.* Libyan J Med, 2025. **20**(1): p. 2528299.
62. Danladi, N.P., et al., *Challenges in Global Distribution and Equitable Access to Monkeypox Vaccines.* Viruses, 2024. **16**(12).
63. Sulaiman, S.K., et al., *Global prevalence and correlates of mpox vaccine acceptance and uptake: a systematic review and meta-analysis.* Commun Med (Lond), 2024. **4**(1): p. 136.
64. Du, M., et al., *Mpox vaccination hesitancy, previous immunisation coverage, and vaccination readiness in the African region: a multinational survey.* EClinicalMedicine, 2025. **80**: p. 103047.
65. Zheng, M., et al., *Mpox Vaccination Hesitancy and Its Associated Factors among Men Who Have Sex with Men in China: A National Observational Study.* Vaccines (Basel), 2023. **11**(9).
66. Bogacka, A., et al., *Mpox unveiled: Global epidemiology, treatment advances, and prevention strategies.* One Health, 2025. **20**: p. 101030.
67. Callaby, H., et al., *Mpox: current knowledge and understanding-a scoping review.* FEMS Microbiol Rev, 2025. **49**.
68. Huang, C.Y., S.B. Su, and K.T. Chen, *A review of epidemiology, diagnosis, and management of Mpox: The role of One Health.* Glob Health Med, 2025. **7**(1): p. 1-12.

69. Karagoz, A., et al., *Monkeypox (mpox) virus: Classification, origin, transmission, genome organization, antiviral drugs, and molecular diagnosis.* J Infect Public Health, 2023. **16**(4): p. 531-541.
70. *Note from the editors: WHO declares mpox outbreak a public health emergency of international concern.* Euro Surveill, 2024. **29**(33).
71. Bhagat, R., G. Kaur, and P. Seth, *Molecular mechanisms of zika virus pathogenesis: An update.* Indian J Med Res, 2021. **154**(3): p. 433-445.
72. Bernhauerová, V., V.V. Rezelj, and M. Vignuzzi, *Modelling Degradation and Replication Kinetics of the Zika Virus In Vitro Infection.* Viruses, 2020. **12**(5).
73. Rossignol, E.D., et al., *Zika virus induced cellular remodelling.* Cell Microbiol, 2017. **19**(8).
74. Vue, D. and Q. Tang, *Zika Virus Overview: Transmission, Origin, Pathogenesis, Animal Model and Diagnosis.* Zoonoses, 2021. **1**(1).
75. Lu, J., et al., *Mpox (formerly monkeypox): pathogenesis, prevention, and treatment.* Signal Transduct Target Ther, 2023. **8**(1): p. 458.
76. Malik, S., et al., *Monkeypox Virus: A Comprehensive Overview of Viral Pathology, Immune Response, and Antiviral Strategies.* Vaccines (Basel), 2023. **11**(8).
77. Pinto, P., et al., *Mpox Person-to-Person Transmission-Where Have We Got So Far? A Systematic Review.* Viruses, 2023. **15**(5).
78. Yadav, R., et al., *Mpox 2022 to 2025 Update: A Comprehensive Review on Its Complications, Transmission, Diagnosis, and Treatment.* Viruses, 2025. **17**(6).
79. Koike, H. and M. Katsuno, *Emerging infectious diseases, vaccines and Guillain-Barré syndrome.* Clin Exp Neuroimmunol, 2021. **12**(3): p. 165-170.
80. Dirlikov, E., et al., *Acute Zika Virus Infection as a Risk Factor for Guillain-Barré Syndrome in Puerto Rico.* Jama, 2017. **318**(15): p. 1498-1500.
81. Koike, H., A. Chiba, and M. Katsuno, *Emerging Infection, Vaccination, and Guillain-Barré Syndrome: A Review.* Neurol Ther, 2021. **10**(2): p. 523-537.
82. Shan, C., X. Xie, and P.Y. Shi, *Zika Virus Vaccine: Progress and Challenges.* Cell Host Microbe, 2018. **24**(1): p. 12-17.
83. Barouch, D.H., S.J. Thomas, and N.L. Michael, *Prospects for a Zika Virus Vaccine.* Immunity, 2017. **46**(2): p. 176-182.
84. Pattnaik, A., B.R. Sahoo, and A.K. Pattnaik *Current Status of Zika Virus Vaccines: Successes and Challenges*. Vaccines, 2020. **8**, DOI: 10.3390/vaccines8020266.
85. Nazneen, F., et al., *An effective live-attenuated Zika vaccine candidate with a modified 5' untranslated region.* npj Vaccines, 2023. **8**(1): p. 50.
86. Morabito, K.M. and B.S. Graham, *Zika Virus Vaccine Development.* J Infect Dis, 2017. **216**(suppl_10): p. S957-s963.
87. Wang, Y., et al., *Current Advances in Zika Vaccine Development.* Vaccines (Basel), 2022. **10**(11).
88. Abrams, R.P.M., et al., *Therapeutic candidates for the Zika virus identified by a high-throughput screen for Zika protease inhibitors.* Proc Natl Acad Sci U S A, 2020. **117**(49): p. 31365-31375.
89. Chellasamy, S.K. and S. Devarajan, *Identification of Potential Lead Molecules for Zika Envelope Protein from In Silico Perspective.* Avicenna J Med Biotechnol, 2019. **11**(1): p. 94-103.
90. Weidenthaler, H., et al., *Real-world safety data for MVA-BN: Increased frequency of syncope following intradermal administration for immunization against mpox disease.* Vaccine, 2024. **42**(22): p. 126024.
91. Forbes, N., et al., *Summary of the National Advisory Committee on Immunization (NACI) Statement-Updated guidance on Imvamune in the context of a routine immunization program.* Can Commun Dis Rep, 2025. **51**(1): p. 1-6.

92. Sah, R., et al., *Monkeypox (Mpox) vaccines and their side effects: the other side of the coin.* Int J Surg, 2023. **109**(2): p. 215-217.
93. McKenna, S.A., et al., *Uncoupling of RNA binding and PKR kinase activation by viral inhibitor RNAs.* J Mol Biol, 2006. **358**(5): p. 1270-85.
94. Koliński, M., E. Kałużna, and M. Piwecka, *RNA-protein interactomes as invaluable resources to study RNA viruses: Insights from SARS CoV-2 studies.* Wiley Interdiscip Rev RNA, 2022. **13**(6): p. e1727.
95. Schmidt, N., et al., *The SARS-CoV-2 RNA-protein interactome in infected human cells.* Nat Microbiol, 2021. **6**(3): p. 339-353.
96. Jønsson, K.L., et al., *IFI16 is required for DNA sensing in human macrophages by promoting production and function of cGAMP.* Nature Communications, 2017. **8**(1): p. 14391.
97. Orzalli, M.H., et al., *cGAS-mediated stabilization of IFI16 promotes innate signaling during herpes simplex virus infection.* Proc Natl Acad Sci U S A, 2015. **112**(14): p. E1773-81.
98. Labidi, H., et al., *Indications of chemical bond contrast in AFM images of a hydrogen-terminated silicon surface.* Nature Communications, 2017. **8**(1): p. 14222.
99. Arakawa, M. and E. Morita, *Flavivirus Replication Organelle Biogenesis in the Endoplasmic Reticulum: Comparison with Other Single-Stranded Positive-Sense RNA Viruses.* Int J Mol Sci, 2019. **20**(9).
100. Meyer, A., et al., *An RNA Thermometer Activity of the West Nile Virus Genomic 3'-Terminal Stem-Loop Element Modulates Viral Replication Efficiency during Host Switching.* Viruses, 2020. **12**(1).
101. Berzal-Herranz, A., et al., *The Genomic 3' UTR of Flaviviruses Is a Translation Initiation Enhancer.* Int J Mol Sci, 2022. **23**(15).
102. Sanford, T.J., et al., *Circularization of flavivirus genomic RNA inhibits de novo translation initiation.* Nucleic Acids Research, 2019. **47**(18): p. 9789-9802.
103. Fajardo, T., et al., *The flavivirus polymerase NS5 regulates translation of viral genomic RNA.* Nucleic Acids Res, 2020. **48**(9): p. 5081-5093.
104. Zhang, Y., et al., *Replication is the key barrier during the dual-host adaptation of mosquito-borne flaviviruses.* Proc Natl Acad Sci U S A, 2022. **119**(12): p. e2110491119.
105. Slonchak, A., et al., *Structural analysis of 3'UTRs in insect flaviviruses reveals novel determinants of sfRNA biogenesis and provides new insights into flavivirus evolution.* Nat Commun, 2022. **13**(1): p. 1279.
106. Ye, J., et al., *Immune evasion strategies of flaviviruses.* Vaccine, 2013. **31**(3): p. 461-471.
107. Bavia, L., et al., *A glance at subgenomic flavivirus RNAs and microRNAs in flavivirus infections.* Virology Journal, 2016. **13**(1): p. 84.
108. Michalski, D., et al., *Zika virus noncoding sfRNAs sequester multiple host-derived RNA-binding proteins and modulate mRNA decay and splicing during infection.* J Biol Chem, 2019. **294**(44): p. 16282-16296.
109. Mazeaud, C., et al., *Zika virus remodels and hijacks IGF2BP2 ribonucleoprotein complex to promote viral replication organelle biogenesis*. 2024, eLife Sciences Publications, Ltd.
110. Hou, S., et al., *Zika Virus Hijacks Stress Granule Proteins and Modulates the Host Stress Response.* J Virol, 2017. **91**(16).
111. Chen, X., et al., *Zika virus RNA structure controls its unique neurotropism by bipartite binding to Musashi-1.* Nature Communications, 2023. **14**(1): p. 1134.
112. Chen, X., et al., *Zika virus RNA structure controls its unique neurotropism by bipartite binding to Musashi-1.* Nat Commun, 2023. **14**(1): p. 1134.
113. Li, P., et al., *Integrative Analysis of Zika Virus Genome RNA Structure Reveals Critical Determinants of Viral Infectivity.* Cell Host Microbe, 2018. **24**(6): p. 875-886.e5.
114. Merino, V.F., et al., *Nucleolin mediates SARS-CoV-2 replication and viral-induced apoptosis of host cells.* Antiviral Res, 2023. **211**: p. 105550.

115. Sekiya, T., et al., *Assembly and remodeling of viral DNA and RNA replicons regulated by cellular molecular chaperones.* Biophys Rev, 2018. **10**(2): p. 445-452.
116. Cong, R., et al., *Interaction of nucleolin with ribosomal RNA genes and its role in RNA polymerase I transcription.* Nucleic Acids Res, 2012. **40**(19): p. 9441-54.
117. Okuwaki, M., et al., *RNA-recognition motifs and glycine and arginine-rich region cooperatively regulate the nucleolar localization of nucleolin.* J Biochem, 2021. **169**(1): p. 87-100.
118. Masuzawa, T. and T. Oyoshi, *Roles of the RGG Domain and RNA Recognition Motif of Nucleolin in G-Quadruplex Stabilization.* ACS Omega, 2020. **5**(10): p. 5202-5208.
119. Yang, D., *G-Quadruplex DNA and RNA.* Methods Mol Biol, 2019. **2035**: p. 1-24.
120. Lyu, K., et al., *RNA G-quadruplexes (rG4s): genomics and biological functions.* Nucleic Acids Research, 2021. **49**(10): p. 5426-5450.
121. Varshney, D., et al., *The regulation and functions of DNA and RNA G-quadruplexes.* Nat Rev Mol Cell Biol, 2020. **21**(8): p. 459-474.
122. Chen, L., et al., *Structural basis for nucleolin recognition of MYC promoter G-quadruplex.* Science, 2025. **388**(6744): p. eadr1752.
123. Watson, J.D. and F.H.C. Crick, *Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid.* Nature, 1953. **171**(4356): p. 737-738.
124. Bansal, A., S. Kaushik, and S. Kukreti, *Non-canonical DNA structures: Diversity and disease association.* Front Genet, 2022. **13**: p. 959258.
125. Li, Z., et al., *G-quadruplexes in genomes of viruses infecting eukaryotes or prokaryotes are under different selection pressures from hosts.* Journal of Genetics and Genomics, 2022. **49**(1): p. 20-29.
126. Marsico, G., et al., *Whole genome experimental maps of DNA G-quadruplexes in multiple species.* Nucleic Acids Research, 2019. **47**(8): p. 3862-3874.
127. Stewart, H., et al., *Identification of novel RNA secondary structures within the hepatitis C virus genome reveals a cooperative involvement in genome packaging.* Scientific Reports, 2016. **6**(1): p. 22952.
128. Bian, W.X., et al., *Binding of cellular nucleolin with the viral core RNA G-quadruplex structure suppresses HCV replication.* Nucleic Acids Res, 2019. **47**(1): p. 56-68.
129. Cui, H. and L. Zhang, *G-Quadruplexes Are Present in Human Coronaviruses Including SARS-CoV-2.* Frontiers in Microbiology, 2020. **11**.
130. Beaudoin, J.D. and J.P. Perreault, *5'-UTR G-quadruplex structures acting as translational repressors.* Nucleic Acids Res, 2010. **38**(20): p. 7022-36.
131. Gomez, D., et al., *A G-quadruplex structure within the 5'-UTR of TRF2 mRNA represses translation in human cells.* Nucleic Acids Res, 2010. **38**(20): p. 7187-98.
132. Morris, M.J. and S. Basu, *An unusually stable G-quadruplex within the 5'-UTR of the MT3 matrix metalloproteinase mRNA represses translation in eukaryotic cells.* Biochemistry, 2009. **48**(23): p. 5313-9.
133. Meier-Stephenson, V., et al., *Identification and characterization of a G-quadruplex structure in the pre-core promoter region of hepatitis B virus covalently closed circular DNA.* J Biol Chem, 2021. **296**: p. 100589.
134. Ruggiero, E., et al., *G-Quadruplex Targeting in the Fight against Viruses: An Update.* Int J Mol Sci, 2021. **22**(20).
135. Pardi, N. and D. Weissman, *Development of vaccines and antivirals for combating viral pandemics.* Nat Biomed Eng, 2020. **4**(12): p. 1128-1133.

Chapter 2. Applied Methodologies

2.1 Foreword

Chapters 2 involves a brief overview of each of the various techniques used for my experiments for identification and biophysical characterization of viral nucleic acids and the various host proteins that are hijacked by them. The primary techniques I employed are microscale thermophoresis (MST), small-angle X-ray scattering (SAXS), circular dichroism (CD) spectroscopy, and size exclusion chromatography coupled with multi-angle and dynamic light scattering (SEC-MALS). This chapter explains the theory, application, and considerations regarding these techniques. All experiments were performed in-house.

2.2 Microscale thermophoresis

2.3 Theory

The employment of MST for scientific research begins with the Ludwig-Soret effect, discovered in 1856 by Carl Ludwig [1, 2]. This effect, described as thermophoresis, thermomigration, or thermodiffusion, explains thermodiffusion of matter in gaseous states rather than liquid states [3]. MST has allowed for the practical application of the Ludwig-Soret effect enabling the researcher to understand how biomolecular interactions behave when infrared radiation (IR) and/or fluorescence light is applied. Following this the researcher can then apply mathematical expressions to define the behaviors of these interactions and thermodiffusion [4]. Highlighting the instrument we utilized, the NanoTemper Monolith™ MST was used to measure thermodiffusion; this technique can sensitively measure the fluorescence of a labeled species, and if that species is interacting with a ligand that is typically serially diluted across 16 tubes.

There are two factors, according to NanoTemper™ that contribute to fluorescence variation across the tubes – one is the temperature-related intensity change (TRIC), which is a phenomenon that occurs with certain fluorophores. Some fluorophores are sensitive to temperature fluctuations, which in turn cause fluorescence intensity differences that otherwise would be at a specific range when temperature is isothermal. The second factor that NanoTemper™ mentions is the thermodiffusion that takes place on molecules as they move away from the produced heat gradient, with each species having its own specific diffusion characteristics based on shape, mass, charge, and hydration shell. Its also important to highlight that the temperature increase that the IR provides is always < 10°C to prevent non-physiological unfolding or shape/charge landscape changes of the species under study. The mathematics explaining the process of thermodiffusion of the fluorescently labelled molecule over time when the IR laser is applied is defined by:

$$\frac{\partial}{\partial T}(cF) = F\frac{\partial c}{\partial T} + c\frac{\partial F}{\partial T}$$

wherein $c\frac{\partial F}{\partial T}$ explains the change in fluorescence over time due to TRIC, and $F\frac{\partial c}{\partial T}$ gives the change in thermophoretic change in concentration, altogether having the summation of these playing a role in thermomigration in the MST. NanoTemper™ continues by mentioning that during the measurement phase, the fluorescence laser is initially shone onto the capillary for 3 seconds to provide a baseline for steady-state fluorescence, labelled F_0 (Fig. 2.1). Once F_0 phase is complete, the IR laser is then shone onto the capillary in tandem with the fluorescence laser, known as the MST-on time (~20-second time frame). During this moment, the fluorescent components in the tube experience thermodiffusion and move away from the heat source, causing a decrease in the focused fluorescence beam. At the end of the 20 seconds, the fluorescence intensity at that moment is known as F_1 (Fig. 2.1). NanoTemper™ continues by mentioning that as the MST sequentially repeats this process over all 16 tubes, and if an interaction occurs, the thermodiffusion of the fluorescent species will change, indicating a binding event, altogether causing a migration change due to size, charge distribution, and hydration shell. The measured fluorescence values are then normalized by using the equation:

$$F_{norm} = \frac{F_1}{F_0}$$

By taking the F_{norm} at a specific time point of each of the capillaries and plotting it as a function of concentration, a dissociation constant (K_d) can be determined. According to NanoTemper™, if an interaction is occurring, the change in fluorescence across the capillaries must be above the signal-to-noise threshold, which places that value > 5 response units. The noise in MST is defined as the standard deviation between responses of experimental data and fitted data, therefore, for your data to agree that an interaction occurred the response amplitude (difference between the bound and unbound state) must be > 5 response units above the standard deviation noise.

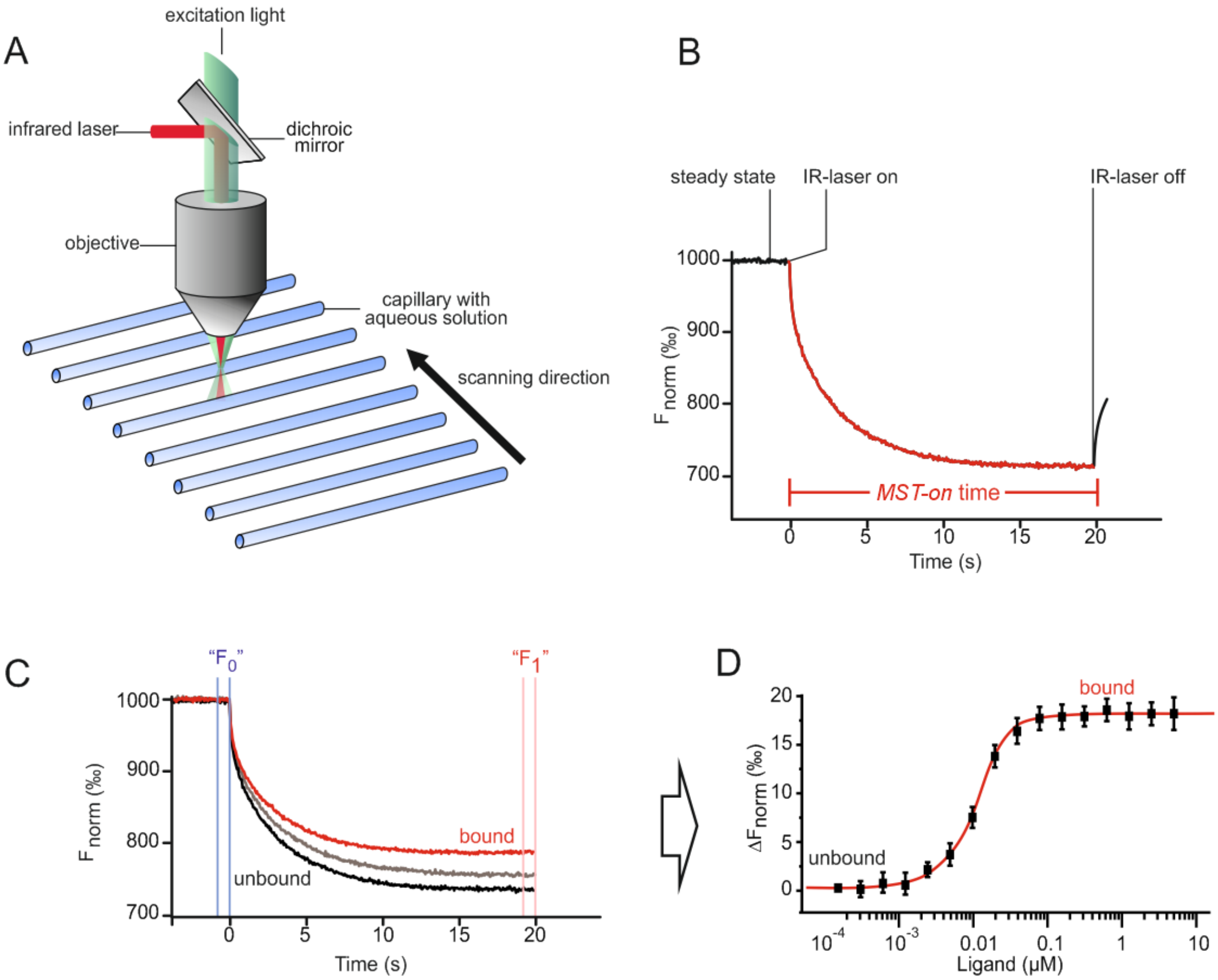


Figure 2.1. MST data collection process. Panels **A**-**D** are schematics of the MST process from the user manual for the NanoTemper Monolith™ MST. Panel **A** shows the physical setup of the experimental setup in the machine with the IR and fluorescent lasers and the dichroic mirror that focuses the photons onto the capillary. Each tube is scanned individually, and the fluorescence collection is represented in Panels **B** & **C**, plotting the normalized fluorescence (F_{norm}) as a function of time (seconds). Panel **B** represents the data collection scan for one capillary in the MST. The MST IR laser on time is ~20 seconds long, followed by a back-diffusion event once the IR laser is off. Panel **C** shows the F_0 and F_1 fluorescence collection from the steady state initial fluorescence to the end of the 20-second data collection. Furthermore, when a binding event occurs, a difference in the thermodiffusion and the fluorescence intensity must occur above the signal-to-noise ratio. Panel **D** represents the change in F_{norm} as a function of ligand concentration, showing the difference in fluorescence of the bound and unbound states. Moreover, the plotted points are fitted using the NanoTemper™ MO.Affinity Analysis v2.3 software, which once fitted, gives the K_d.

The K_d is determined from the final plotted/fitted data, and is calculated from the law of mass action:

$$K_d = \frac{[A][L]}{[AL]}$$

wherein [A] is the concentration of free fluorescent molecule, [L] is the concentration of free ligand, and [AL] is the concentration of the complex of both A & L. MST provides wide versatility in what species or biomolecules you are able to measure with, the fluorophore choice, and allow for an accurate determination of the K_d that is a worthy instrumental addition to other biophysical techniques that researchers use to elucidate kinetics.

2.4 Applications

MST is used in a wide range of chemical and biochemical species, which include protein-protein, RNA-RNA, protein-RNA, protein-DNA, DNA-small molecule, and protein-small molecule [5-12]. Regarding protein-protein interactions, a research group tested this using the beta-lactamase TEM1 with the beta-lactamase TEM1 inhibitory protein (BLIP). Using purified wild-type proteins, they achieved a K_d of 3.8 nM. The researchers also tested the interaction from cell lysate, and they achieved a K_d of 10 nM. It's worth noting that this specific research group was able to characterize and show the interaction in a purified state with the two components in the capillary as well as in cell lysate, however, this is not the case for all biomolecular interaction studies. Some interactions might require various conditions or buffers to determine a K_d. Moreover, this answer could not be elucidated using isolated protein-small molecule interactions [13]. This can be shown with a different research group revealing that the cAMP-dependent kinase PKA inhibitor, quercetin, had an affinity of 130 nM in a two-component mixture, but with 5% human serum, the affinity jumped to 6 µM, and in 30% serum, 50 µM, indicating that not all interaction studies can be performed in a multi-variable, heterogeneic mixture [13].

Regarding the other aforementioned interactions, our lab has shown protein-RNA, protein-DNA, DNA-small molecule, and RNA-RNA interactions by performing all of our studies in two-component conditions instead of multi-variable heterogenetic mixtures such as lysate or serum. We have demonstrated in two studies that DDX3X and DDX17 bind and unwind the JEV/ZIKV 5' terminal region (TR), and Rift Valley Fever virus non-coding RNA, respectively [14, 15]. For the DDX3X-JEV/ZIKV 5' TR, a K_d was determined at 1.66 and 7.06 µM respectively [14, 15]. Moreover, for DDX17, it showed 5.78 µM for the intergenic region and 9.85 µM for the 5' non-coding region [14, 15]. It is worth addressing that the MST can be used for other investigations besides a K_d, as seen in the previously mentioned discoveries by our research group of DDX3X and DDX17 binding to JEV, ZIKV, and Rift Valley Fever virus [14, 15]. We demonstrated using the binding check experiment on the NanoTemper Monolith™, and determined that these helicases unwound these RNAs when ATP was titrated in, along with a labelled oligonucleotide used to bind to a complementary region if the RNA was unfolded [14, 15].

Furthermore, for protein-DNA interactions, we discovered that the human G4 binding protein, DHX36, binds to Mpox virus DNA G4s (see Chapter 8). Our work identified two regions within the Mpox genome that contain G4s, and we elucidated a K_d of the first quadruplex, denoted MP1, at 44 nM, and MP2 at 611 nM, showing that differences in quadruplex structure may be playing a role in the kinetic differences shown with DHX36. This can be seen in other literature supporting DHX36 having binding/unfolding variation depending on the spacing between guanine pairs, and the types of nucleotides present between the guanine pairs [16, 17]. Along with the research we published on Mpox, we further discovered with MST that the G4-binding molecule, TMPyP4, bound to the MP1 and MP2 G4s [10], revealing that MST can measure DNA-small-molecule interactions. Finally, our lab demonstrated that the MST could determine the K_d of RNA-RNA interactions, which was shown in the JEV 5' and 3' TRs, with a K_d determined at 169 nM [18].

Collectively, these examples all provide compelling evidence that MST is a valuable tool for studying interactions and potential enzymatic activity from a wide range of biological and small molecules.

2.5 Considerations

MST has advantages for studying interactions, but each tool used in research carries limitations/caveats. The first limitation is generally using any instruments that utilize fluorescence-based detection methods. With any fluorophore tagged onto a molecule, certain chemistry/electrostatic influences can cause quenching. This quenching can lower fluorescence intensity, and consequently, the need to increase the concentration of the species to get a high enough signal to measure the rate of diffusion, which can altogether hinder the accurate determination of a K_d. Secondly, the fluorophore can cause an alteration in binding properties or cause steric hindrance, as fluorophores tend to be flat and hydrophobic. Furthermore, it does not provide information on other kinetic constants such as the k_{on} and k_{off}, but solely K_d. With these present caveats, it is imperative that when you use MST you complement the results with another technique that can validate binding, especially when determining dissociation constants. Lastly, MST cannot convey any information on the number of species that are binding to one substrate, and/or, where they are binding. For example, Chapter 5 discusses a human protein called nucleolin, and we show that it binds to the ZIKV TRs, however, we cannot determine with this technique that multiple nucleolin proteins bound to the viral RNA, or where they bound specifically. Other structural techniques such as X-ray crystallography, cryogenic electron microscopy, or small angle X-ray scattering could complement the MST results and give us insight as to where and how many nucleolin proteins are binding.

2.6 Small-angle X-ray scattering

2.7 Theory

SAXS allows the determination of the overall shape, conformation, and assembly of various monomeric or oligomeric states of proteins, nucleic acids, and their complexes [1]. This technique was invented by André Guinier in 1938 and can resolve particles ranging from 1-200 nm [1]. SAXS instruments use high-intensity X-ray radiation that can be generated by a synchrotron, or an in-house light source, that both bombard a biological sample with a collimated radiation beam [3]. The synchrotron approach is more advantageous for studying biomolecules, as it provides a larger amount of X-ray flux that cannot be achieved with an in-house machine [19]. The higher flux allows a more detailed structure determination of your species [19]. For our purposes, we use the Diamond Light Source synchrotron in Oxford, England to retrieve our scattering data. Once the sample is injected into the chromatography instrument at the facility and separation of heterogeneic species occurs, the species are then passed through the X-ray scattering flow cell and bombarded with the monochromatic X-rays with a wavelength of ~0.7-1.7 Å [20]. During this event, elastic scattering occurs in which an incident photon interacts with atoms throughout the species, followed by release of that energy in the same wavelength of the incident photon and scatters at a specific angle [21]. Many other factors are involved in this event that are difficult to fully quantify, but for the purpose of our research goals with SAXS, light that deviates at low angles (0.1-10 degrees) from the incident photon trajectory is collected using a hybrid-photon-counting detector [22]. The scattering profile collected at this angle provides structural features that can only be visually represented after performing a series of data analyses that will be discussed further.

With the Brownian motion of the biological particles occurring as they move past the X-ray source, the current determined resolution of the generated solution envelopes today is between ~10-1000 Å [2]. This provides a representation of the biomolecules' overall conformation and shape in solution [2]. It also allows for the quantitative and qualitative characterization of flexible or intrinsically disordered proteins [2]. SAXS can also be used to elucidate the intrinsic disorder of proteins or nucleic acids or used as a complementary technique to provide physiological relevance to biomolecules or their complexes [23].

1-dimensional data analysis (Scattering, Guinier, Kratky, Pair-wise distance distribution function plots

There are various software programs that can be utilized to help with analyzing SAXS data – ScÅtter, bioXtas, and ATSAS [24-26]. We primarily use the ATSAS software in our research group.

The first step with plotting the raw data collected is to perform buffer subtraction. Along with the scattering that occurs from the species under study, the buffer components also scatter. The ATSAS program will subtract the buffer and allow only the scattering data from the species under study. After this, the data can now be inputted onto a scattering curve wherein the scattering intensity is plotted as a function of the scattering vector (Fig. 2.2). This scattering plot is then used to plot a Guinier analysis, invented by André Guinier [27]. The reasoning for this analysis is to extract a radius of gyration (R_g) value, which provides details on the size and shape of the species under study, and it can be used to further calculate the molecular weight of the species [26]. This value is calculated based on the root mean square distance of each of the electrons of the species from its center of mass (Fig. 2.2). Next, a Kratky plot is generated, and provides information on the general fold, shape, flexibility, or disorder of the species (Fig. 2.2) [28]. This is performed by plotting the scattering intensity as a function of the square root of the scattering vector multiplied by the determined center of mass from the sample, which was altogether determined in the Guinier plot (Fig. 2.2) [9]. Lastly, for 1-D analysis, we determine the maximum distance of the species under study, called the D_{max}. This is a probabilistic calculation that is determined by developing a pair-wise distance distribution function (P(r)) [26, 29]. This plot is computed using a Fourier transform of the scattering curve and gives the probability of finding two atoms at any given position within the determined structure in space (Fig. 2.2). The tail end of the computed Gaussian curve after Fourier transformation will give the D_{max} value.

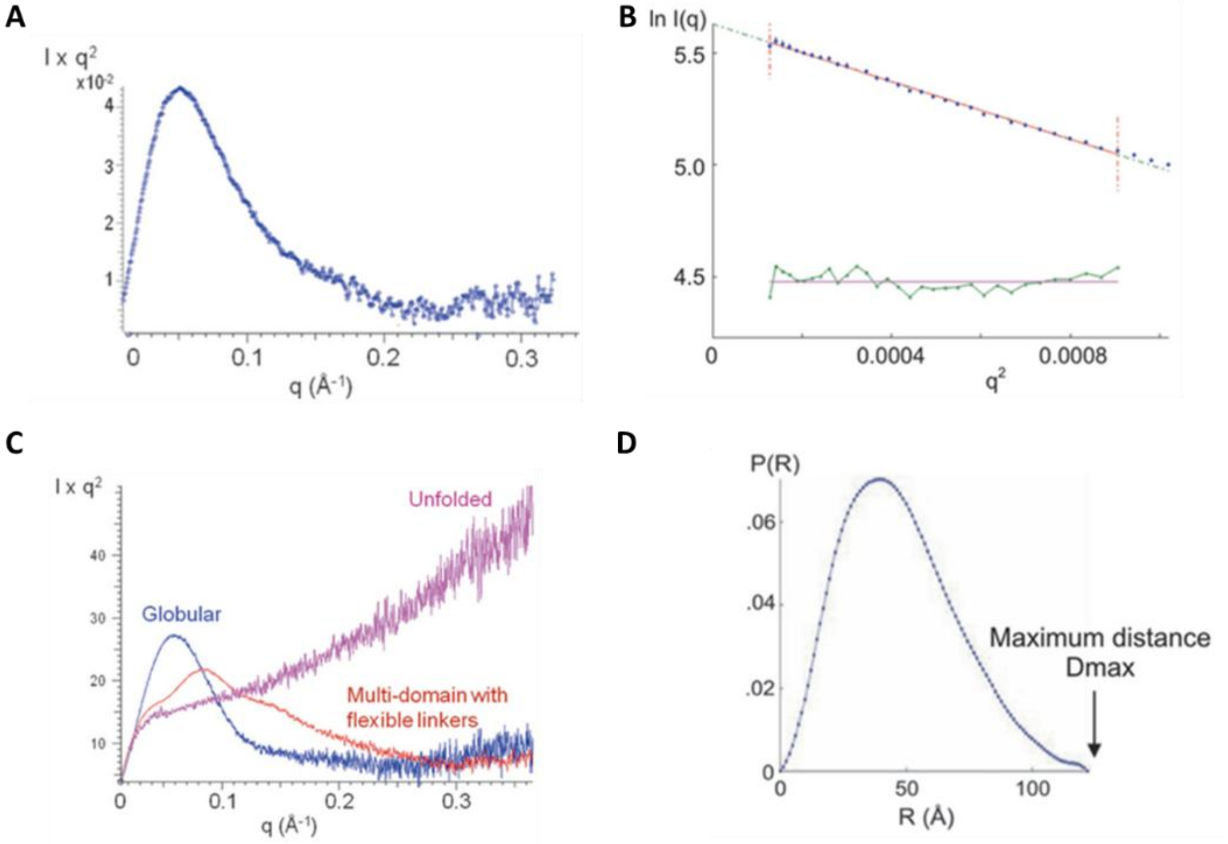


Figure 2.2. SAXS data plots of the 1-D analysis – Panel **A** the scattering curve, Panel **B** the Guinier plot, Panel **C** the Kratky analysis, and Panel **D** the pair-wise distance distribution function, P(r) [30]. these plots are taken from a paper showing the SAXS structure determination of the Dengue virus NS5 RNA-dependent RNA polymerase [30].

Ab initio DAMMIN/DAMMIF model refinement, alignment, and generation

The *Ab initio* modelling that takes 1-D data and converts it to real space data in 3 dimensions (3-D) uses the raw scattering curve of the scattering intensity plotted as a function of the scattering vector [31]. Using the DAMMIN/DAMMIF on the ATSAS suite, researchers can obtain a low-resolution 3-D structure of the species under study [32, 33]. This begins by performing single-phase dummy atom modelling, wherein the data is represented as a collection of interconnected beads, with each bead being a representation of each atom [16,17]. Next, this model is taken into a simulated annealing algorithm that undergoes several repeated annealing steps (~100 in our research group) to approach the desired level of accuracy [31]. This level is determined by the convergence of the generated models, and the chi-square goodness-of-fit test to the experimental data [31]. After refinement, a model ensemble is generated which is then inputted into DAMSEL/DAMCLUST for alignment and clustering [31, 34]. This compares all the models generated using DAMMIN/DAMMIF and can identify the probable models over the outlying ones via superposition. Once the alignment is completed, a set of statistically significant models is selected and classified into clusters of probable structures that align the most with each other [31]. Finally, the software DAMAVER selects the most probable structure(s) based on the clustering that has the highest statistical significance, and the final model is then exported to Pymol for visualization [35]. The models generated from this can further be used for alignment with higher resolution structures that have been determined by a high-resolution instrument such as cryogenic electron microscopy (cryo-EM), X-ray crystallography, and/or regarding RNA research, computational means after secondary structure retrieval using selective 2’-hydroxyl acylation analyzed by primer extension (SHAPE).

2.8 Applications

Cryo-EM, X-ray crystallography, and/or in the case of RNA, SHAPE, subjects biomolecules into non-physiological states for data retrieval [36, 37]. Hence, the structures determined using these techniques may not reflect the relevant physiological structure of them. Furthermore, these techniques require high-purity samples and long data processing times. Conversely, SAXS enters through a chromatography step prior to data collection and allows researchers to study biomolecules under near-physiological conditions in contrast with high-resolution techniques.

Regarding RNA, the low-resolution SAXS structures can be used as constraints for developing high-resolution structures of RNAs by integrating the secondary structure information derived from SHAPE. For this purpose, the SHAPE data is subjected to computational modelling to generate a pool of possible high-resolution conformations for the given two-dimensional structure [20]. These models can be further improved using thermodynamic constraints to provide a physiologically relevant model in a relatively less time [20].

Our lab has biophysically characterized many viral nucleic acids and human protein structures using SAXS, and currently, we are currently pursuing in-house computational modelling with our determined RNA structures and fitting them with SHAPE data. As an example of our lab's SAXS structure determination, we have published data on the pre-core G-quadruplex (G4) of the hepatitis B virus, a non-canonical structure that the virus requires for viral replication [22]. We compared the wild-type pre-core G4 with three mutant constructs showing that the wild-type structure is the most compact and folded (Fig. 2.3). Hepatitis B virus is a hepatovirulent virus that chronically infects 250 million individuals worldwide, and infection of this virus significantly increases the risk of developing liver cirrhosis, hepatic failure, and hepatocellular carcinoma [38]. Using this method as a guideline our lab is currently attempting to study G4s from untranslated regions (UTR) of flaviviruses. Further G4 characterization of the Mpox virus will be discussed in chapter 8.

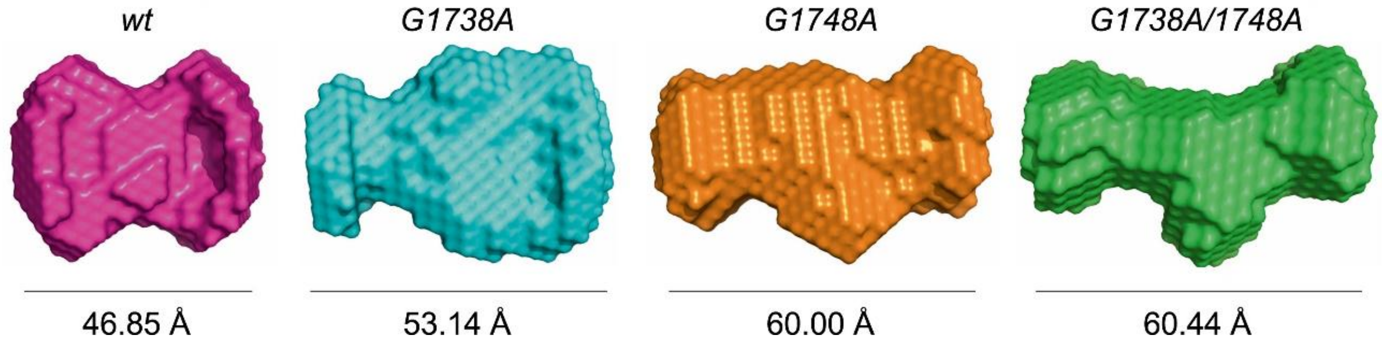


Figure 2.3. Averaged-filtered models derived using DAMMIN and DAMAVER of the Hepatitis B virus pre-core G4 along with a comparison of various mutants of the wild-type pre-core G4. The values below are a comparison of the D_{max}, which indicate that the wild-type structure is the most compact/folded in contrast with the other mutants, suggesting that a G4 is formed. For additional details, see [38].

2.9 Considerations

There are several potential variables to consider when performing SAXS. Firstly, SAXS requires a homogeneous, monodisperse, aggregation-free sample to have accurate data collection [37]. Failure to achieve this would result in a poor fit for the data analysis, and time lost in experimentation [52]. Consequently, SAXS requires potential optimization of overall biomolecule design, and buffer and concentration optimization [37]. Concentration also plays a role, as SAXS requires a small amount of sample, but at a higher concentration, which can lead to concentration-dependent interactions that may reflect physiologically relevance [52]. Secondly, radiation damage can occur, therefore, careful optimization of flux, as well as exposure time is needed for data collection [53, 54]. Lastly, SAXS produces physiological relevance over other structural determination techniques, albeit with low resolution [55]. Because of this, the orientation of determined structures can be difficult to decipher if the species under study has any symmetrical resemblance [38-40]. Ultimately, SAXS is a complementary technique for understanding the general solution structure and size of your species under study [56]. If a researcher requires higher resolution, for example, drug testing or general binding/targeting, SAXS provides a preliminary picture [41]. For our purposes in the lab, SAXS provides this preliminary result, but for our research group to perform further structure detail studies we require higher resolution techniques to provide a comprehensive view of the species under study. For any targeting of ZIKV or Mpox replication within our studies, further structural determination is paramount (chapters 5, 6, and 8).

2.10 Circular Dichroism Spectroscopy

2.11 Theory

Circular dichroism (CD) spectroscopy is a structural characterization technique that utilizes circular polarized ultraviolet light and exposes it to a chemical or biological species to gain insights into specific characteristics or secondary structures [57]. Structural biologists utilize this technique on various biomolecules such as nucleic acids, proteins, lipids, and sugars, altogether at a wide range of sizes and complexes [57, 58]. Optically active chiral centers within the biomolecules under study will absorb one of the circularly polarized lights (ie. either clockwise or counterclockwise), while the other travels past the sample [57]. Rotationally polarized light highly depends on the molecular conformation and chirality to retrieve a specific signature, and regarding biological molecules, specific secondary structures will have consistent absorption signatures, which allow researchers to elucidate if their species under study contain specific structures or folding profiles [57]. Furthermore, optimal data collection relies on the concentration of your species and the secondary structure quantity, which consequently, requires the researcher to fine tune the concentration for optimal signal and results [57]. CD spec, albeit not a high-resolution technique, is a non-destructive approach that allows for the elucidation of physiologically relevant structural information without the invasiveness of high energy photon destruction, or fluorescent labeling [59]. Most importantly, you can retrieve your samples after data collection for any downstream experiments [60].

The CD spec instrument produces ultraviolet (UV) light that is generated by a short arc xenon lamp which is constructed with specialized high-purity synthetic fused silica, or fused quartz [5]. This lamp produces wavelengths from 175-800 nanometers, with biological applications occurring on average from 180-400 nm [61]. This select wavelength range tailored for biomolecules is due to their chiral centers absorbing circularly polarized light within that specified and optimized wavelength [61]. Any wavelength outside of that range is used for other applications not related to elucidating biomolecular secondary structure information [5]. Furthermore, the UV photons generated by the xenon lamp travel through a polarizer filter that only allows photons in a specific plane to pass through [61]. These photons then encounter a photoelastic modulator made of quartz, which is used to convert the linearly polarized light to circularly polarized light [61]. When a 50 kHz frequency of light is attained, the quartz piece is stress-induced, causing circular birefringence which releases clockwise and counterclockwise circularly polarized light [61]. This light travels to the sample cuvette containing the species under study [5]. The incident photons will pass through the solution, and any optically active molecules will absorb varying amounts of clockwise or counterclockwise polarized light [61]. The bidirectional polarized light is then picked up by the detector, and the molar ellipticity is computed, which is the angular difference at tan(Θ) of the circular incidence photons divided by the concentration of the species and the path length used [62]. The intensity of light detected is proportional to the square of the electric-field vector, and the ellipticity expresses as Θ (radians) [62].

Regarding the characterization of biomolecules that our group researches, there are several secondary structures in proteins and nucleic acids that can be identified, which also include non-canonical structures in nucleic acids, each giving a specific ellipticity profile [63]. Moreover, the CD spec profile produced is dependent on the length and total amount of the secondary structure within the biomolecule [7]. If the biomolecule under study possesses multiple secondary structures, there are deconvolution software's available, such as DichroWeb, BeStSel, CDPro, DichroCalc, or K2D2 for protein species, and NACDDB, CDNuSS, DichroWeb, and CDtoolX for nucleic acids [60, 63-67]. Ultimately, these tools are successful at predicting the secondary structure of biomolecules (see Chapters 5 and 8). However, some considerations must be addressed when using this technique for the retrieval of accurate and precise information.

2.12 Applications

There is previous literature showing the reliability of CD spec. An example can be seen with a research group utilizing streptavidin, gammas B crystallin and s-crystalline C, Avidin, Immunoglobulin G, beta-lactoglobulin, glycogen phosphorylase-b, 3-dehydrogenate, glutamate dehydrogenase I, aldolase, and phosphatidylcholine 2-acylhydrolase, and compared the percentage of secondary structure with determined by high-resolution structures to the determined secondary structure from a CD spec profiles (Fig. 2.6) [68]. Using a goodness-of-fit parameter known as the normalized root mean squared deviation, the structure percentage on CD spec was statistically significant for all structures when compared with the determined crystal structures [68]. This study indicates that CD spec is a powerful tool, as it shows accuracy in elucidating the near-physiological secondary structure of proteins compared to the non-physiological states present in X-ray crystallography. Because of this revelation, CD spec could be used to elucidate the differences that could be present between a crystal structure and the conformational differences a protein may have in solution in a more physiological state.

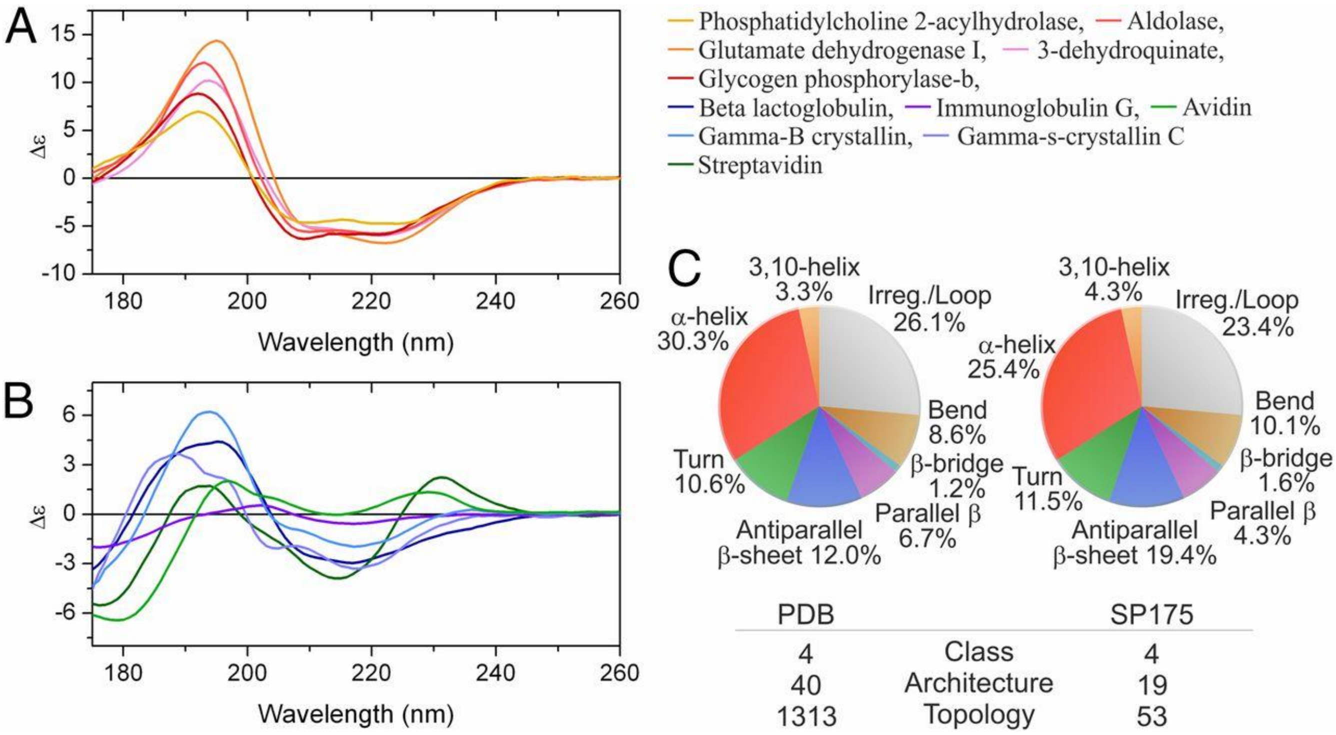


Figure 2.6. Panels **A** and **B** represent CD spec profiles with the change in ellipticity as a function of wavelength for alpha helix-rich and bet sheet-rich proteins respectively [12]. Panel **C** compares the protein secondary structure percentages of PDB structures (left) and the deconvoluted software's result of the CD spec profiles (right) [12].

An example regarding nucleic acids can be found in a study that developed software that can identify various secondary structures of nucleic acids [63]. This study shows the development of a software known as CD-NuSS (Fig. 2.7). It’s worth noting that nucleic acid structure elucidation is challenging, as the current techniques that can be used (aside from CD spec) are crystallography, NMR, cryogenic electron microscopy, or SHAPE [7]. Most of these techniques require expensive instruments and infrastructure, as well as heavy computation to elucidate structure. Furthermore, aside from NMR, these techniques subject the RNA to non-physiological conditions/states, which can make it difficult to demonstrate physiological relevance to these structures alone without complementary physiologically relevant techniques (i.e., CD spec, SAXS, analytical ultracentrifugation (AUC), or size exclusion chromatography-multi-angle light scattering (SEC-MALS)) [8]. Moreover, CD spec is limited to the number of varying secondary structures in nucleic acids, as there are no programs that can deconvolute multiple secondary structure RNA species, which will be further discussed in the considerations [8]. Aside from these caveats, they mention that CD spec is a useful tool for elucidating RNA secondary structure without the need for probes, modification, or non-physiological temperatures (dependent on the organism’s living conditions) [7]. The research group further showed that they achieved an 85% success rate of RNA structure identification when compared to structures that have been confirmed by other methods [63]. CD-NuSS is able to predict stem-loops of both RNA and DNA, DNA-RNA duplexes, G-quadruplexes in various forms (parallel/antiparallel, hybrid, iG, G-triplex, parallel/antiparallel triplex, and pentaplex, as well as B, Z, and A form RNA and DNA, as well as i-motifs [63] (Fig. 2.7). Unfortunately, it cannot predict pseudoknots, or reveal specifics of length or structure of stem-loops [63].

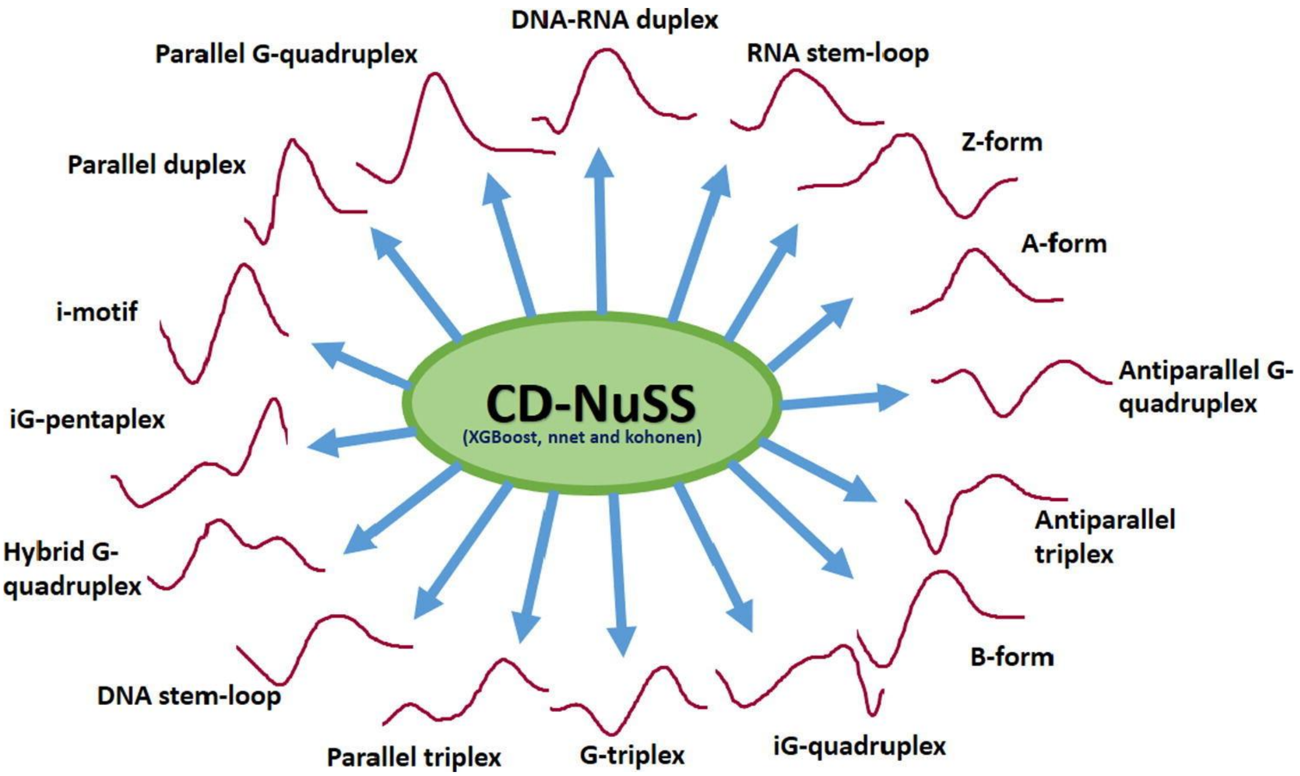


Figure 2.7. The various nucleic acid structures the CD-NuSS software can predict, as well as a schematic of the CD spec ellipticity profile determined for each structure [7]. Furthermore, the green center indicates the machine learning and analysis algorithms used to analyze and determine the structures within the sample investigated [7].

2.13 Considerations

CD spec provides a clear picture of your biomolecules' secondary structure landscape, albeit with a few restrictions or considerations. Firstly, there are buffer restrictions on compounds that contain optically active chiral centres [4]. This can contribute to noise with the CD spec signal [60]. Even when buffer cancellation is performed, optically active buffers can contribute a substantial amount of signal [4]. The threshold for buffer contribution during pre-scan must be <1 absorbance unit [60]. If the researcher approaches beyond the threshold, the data is too noisy. Common buffers that have been reported to be optimal for CD spec include phosphate, Tris, perchlorate, or borate at lower concentrations [60]. Secondly, although data collection remains successful in determining the secondary structure composition/profile through deconvoluted software regarding proteins, this is not the case with nucleic acids [8]. Despite the determination of singular DNA and RNA secondary structure, any multi-secondary structure nucleic acid cannot be determined, as there are no available deconvolution methods to decipher these species [7,8]. Currently, structure determination can only be performed with smaller nucleic acid polymers with single structure topology [7,8]. Furthermore, it can be difficult to obtain structural homogeneity with certain RNAs, as they can possess differences in their overall structure, yet contain the same RNA sequence [69]. Finally, it is important to note that CD spec is a complementary technique and does not extrapolate detailed structural data of a biomolecule, but more so a synergistic technique when coupled with high-resolution techniques mentioned earlier, or SAXS, AUC, and/or SEC-MALS. Ultimately, CD spec is a powerful technique to reveal near-physiological secondary structure profiles of biomolecules across a wide range of organisms that may require salt or temperature accommodations to provide physiological relevance and validation to structure.

2.14 Multi-angle and dynamic light scattering (MALS/DLS)

2.15 Theory

Size exclusion chromatography – multi-angle light scattering (SEC-MALS) is a biophysical technique wherein a biomolecule's physical properties can be determined in solution in near-physiological states. With this technique, researchers can determine the stoichiometry, molecular weights (MWs), oligomerization, and solution size, otherwise known as the radius of hydration (R_h), of the molecule(s) under study [1]. SEC-MALS utilizes a collimated, single-frequency, polarized light beam that illuminates a sample passing through the incident beam [1]. The light interacts with electrons within the biomolecule causing Rayleigh scattering, which consequently releases the photons at various angles and intensities depending on the physical properties of the species. The specific angles/intensities that the sensors detect, as well as the scattering fluctuations that occur over time during dynamic light scattering data collection, reveal details about the MW(s) and the R_h (Stokes radius). Furthermore, based on the UV absorption profiles of complexes and individual species separated by size exclusion coupled with the determined MW, stoichiometry can be assumed. Using the Wyatt technology™ software, we can reveal these structural and/or interaction details of proteins, nucleic acids, polymers, nanoparticles, liposomes, and extracellular vesicles.

During an SEC-MALS run, an HPLC machine is used as the tool for providing a high-pressure flow through the SEC column, as well as the multi-angle and dynamic light scattering instruments [70]. The HPLC contains a specialized SEC column that is specific to a specific size range and shape of the biomolecules under study. This column is imperative for biophysical characterization purposes, as it separates contamination, oligomerization, and bound complexes/unbound species, which altogether promote more refined monodispersity through the entire data collection process in the scattering instruments [70]. Regarding the MALS/DLS instruments from Wyatt Technologies™, they mention that the HPLC machine is calibrated to the scattering instruments to aid in seamless data collection, and the MW determination using MALS is independent of the elution time and does not require elution calibration for assisting in determination. MALS also possesses the ability to determine the MWs of biomolecules without the need for standards to be used.

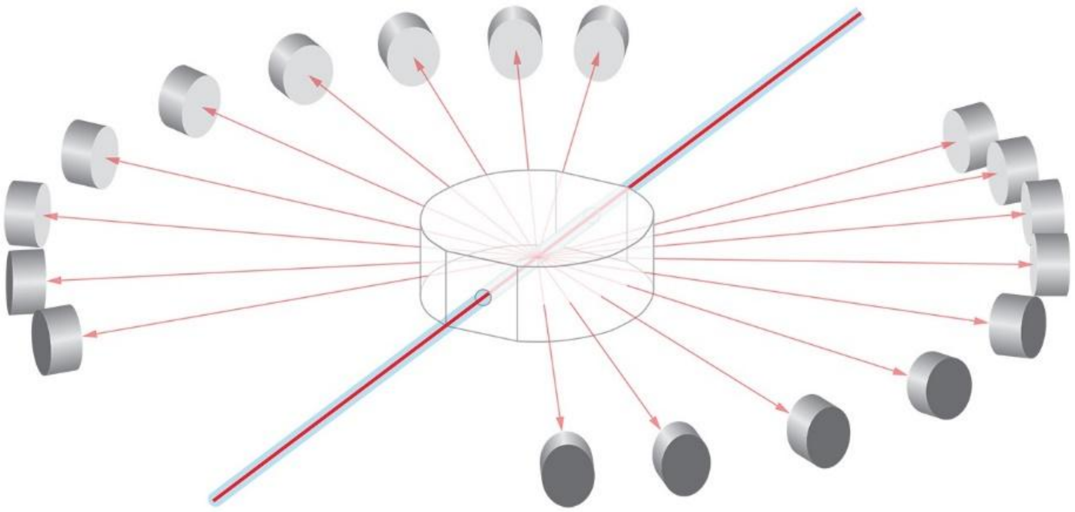


Figure 2.8. A schematic representation of the MALS instrument operation. The sample under study passes through the incident light beam, Rayleigh scattering occurs, and the reflected light at varying intensities is collected at ‘static’ angles which is then used to calculate the MW of the species. Image retrieved from Wyatt Technologies™.

Multi-angle light scattering data collection

MALS, or static light scattering is a biophysical technique wherein monodisperse samples are passed through a collimated laser, and as the light strikes the sample, the light oscillates electrons which then release the light at various angles specific to the position the electron is in on the biomolecule [70]. The intensity of the scattered light on the static positioned detectors is assessed through the Wyatt Technology™ software as it passes through the light source, and a MW can be calculated from this. Furthermore, this can only be determined if the intensity, concentration of species, and $\left(\frac{dn}{dc}\right)$ value is known, which can all be collectively expressed as:

$$I_{scattered} \propto M \cdot c \cdot \left(\frac{dn}{dc}\right)^2$$

wherein the intensity of the scattering light ($I_{scattered}$) is proportional to the molar mass (M) multiplied by the concentration of the species, and $\left(\frac{dn}{dc}\right)^2$ which is a sample-specific value of the change in the refractive index in relation to the change in concentration of the species under study. The $\left(\frac{dn}{dc}\right)$ value is determined by a few methods – the simplest method is to look through previous peer-reviewed literature for the values various research groups used for the biological species, whether its DNA, RNA, or protein. The second method is to perform online measurement, wherein you use a liquid chromatography machine and measure the refractive index on the peaks of your species [71]. However, this method, though robust in determining $\left(\frac{dn}{dc}\right)$, requires expensive chromatography equipment, which can be outside of a researcher's budget. Lastly, a batch approach can be performed but requires an exact concentration of the species, a well-calibrated SEC column, as well as a known $\left(\frac{dn}{dc}\right)$ value [72]. Aside from the cost and optimizing, this stringent approach allows a more detailed $\left(\frac{dn}{dc}\right)$ value when extracting the refractive index.

The type of scattered light collected on the static detectors is primarily Rayleigh scattered light, wherein a released photon and its energy is lower than the incident photon energy prior to absorption [73]. The process of determining this entire process, along with being able to determine the MW and R_h was developed over the past 180 years by many researchers [74-82]. All these data were summated by a researcher known as Bruno Zimm for the determination of MW, which can be expressed as:

$$\frac{K^*c}{R(\theta,c)} = \frac{1}{M_W P(\theta)} + 2A_2c$$

Where K is the stretch factor used in plotting the scattering angles that determine the MW, c is concentration, which is altogether over the $R(\theta,c)$, which is a function of the scattering angle (θ) and the concentration of the species [81, 83]. This equation is equal to 1 over the MW of the species multiplied by $p(\theta)$, which is the angular dependence of the scattered light, plus 2 times the second virial coefficient (A_2), which deals with the osmotic pressure of the system, multiplied by the concentration of the species (c).

Dynamic light scattering (DLS)

In successive order with the SEC-MALS machine, the sample passes from the MALS to the DLS instrument. The difference between the MALS and DLS is that DLS does not measure the scattering angle intensity, but the light intensity fluctuations occur over time as the sample passes through the light source [84, 85]. Using Einstein's expression describing the random motion of particles moving through a solution experiencing Brownian motion, we can determine the radius of hydration by the Stokes-Einstein equation, which is:

$$R_h = \frac{kT}{6\pi\eta D_t}$$

Where η is the intrinsic viscosity, R_h is the radius of hydration, D_t is the diffusion coefficient, k is the Boltzmann constant, and T for temperature [86, 87]. As the particles experience Brownian motion while passing through the light source, they scatter different quantities of light, and the rate of diffusion corresponds to the rate of light fluctuations over time. This time is measured between microseconds to milliseconds, depending on the species being measured [88]. The rate of diffusion occurring, the intrinsic viscosity of the buffer, as well as the light fluctuation of light scattering, we can determine the hydrodynamic radii (R_h) of each homogenous species as they pass through the light source [86-88]. The determined R_h is the computed apparent size of the hydrated/solvated particle, and dynamic, meaning the species are moving/tumbling in solution.

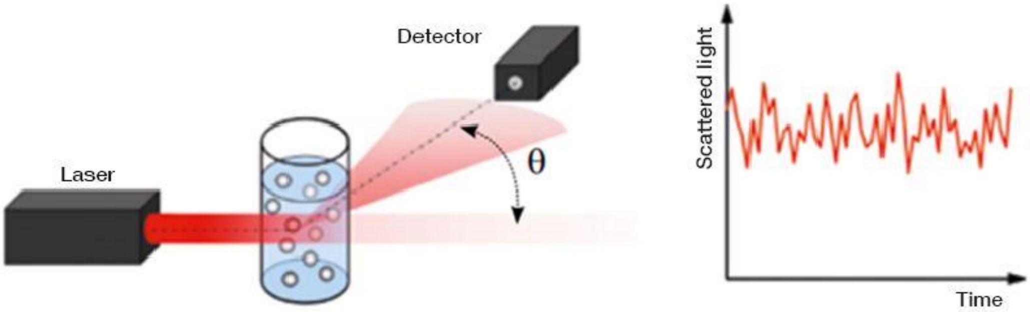


Figure 2.9. A generalized schematic representation of the DLS instrument and its data collection procedure. The detector of the instrument collects at a static angle (left), and the scattering fluctuations (right) that it collects give insight into the general size of the species in solution (R_h). image retrieved from Anton Paar GmbH©.

In summary, SEC-MALS allows the researcher to determine what the hydrodynamic properties are of the species under study, including homo and heterogeneous species. This application provides a strong tool in the pursuit and elucidation of the biophysical properties of biomolecules' behaviour in solution that other instruments that subject molecules to non-physiological states cannot achieve.

2.16 Applications

There are many applicable examples of biophysical characterization on various biomolecules using MALS and DLS instruments, including proteins, nucleic acids, polymers, nanoparticles, liposomes, and extracellular vesicles [89-94]. This instrument can also reveal details on interactions of said biomolecules such as protein-protein, protein-nucleic acid, or nucleic acid-nucleic acid interactions [7, 70, 84, 85, 90, 95].

Many studies are showing these interactions, including some published by our research group, as well as examples provided by Wyatt Technologies™[7, 70, 95-100]. Wyatt has performed refractive index and MALS data collection on various biomolecules, such as protein, DNA, and RNA. Regarding protein characterization, Wyatt was able to identify the MWs of ferritin, aldolase, ovalbumin, and RNase proteins [70]. The determined MWs were further validated using analytical ultracentrifugation (AUC), which showed the same result. The group then showed oligomerization of the protein dimer tubulin, and determined the MWs of the complex, as well as the monomers, again, validating the MWs by AUC [70]. Furthermore, another example can be seen with adeno-associated virus (AAV) vectors and characterizing the encapsulating efficiency of DNA into the AAV [95]. The research group was able to characterize and determine the MW, R_h, and concentrations of loaded and empty vectors after a loading reaction was performed. They also showed the loading and molecular weight variations in the presence of varying pHs and salt concentrations [95]. This evidence shows the effectiveness of using MALS/DLS, even in the presence of an oligomeric protein with a nucleic acid polymer encapsulated inside.

Some examples of the characterization of nucleic acids are seen within our research group, specifically, the characterization of the human long- non-coding RNA (lncRNA), LincRNA-p21 [51]. This lncRNA is a transcriptional repressor in the p53 pathway and plays a role in triggering apoptosis. The research set out to investigate the structure of the sense and anti-sense LincRNA-p21, as these structures have not been determined. We utilized MALS/DLS to determine if the RNA is monomeric and monodisperse, and from the results below in Fig. 2.10, the MW determined by SEC-MALS agrees with the theoretical MW [51]. We also determined that the sense and anti-sense RNA displays asymmetrical, extended structures *in vitro*, in near-physiological solution conditions (Fig. 2.10).

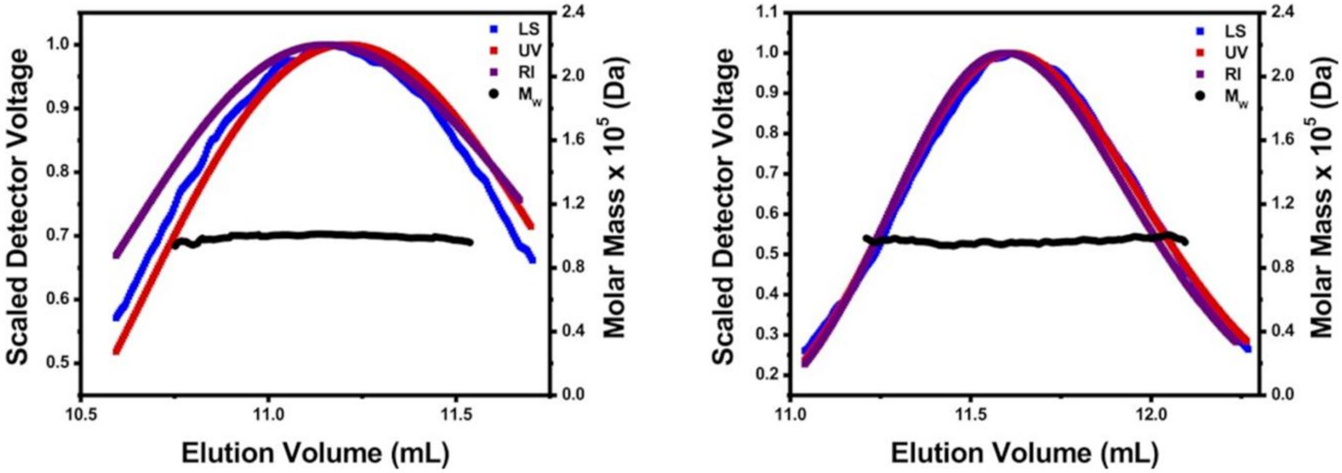


Figure 2.10. SEC-MALS results show the determined MW of the sense (right) and antisense (left) LincRNA-p21 AluSx1. The calculated MWs were both 99 and 94 kDa for the sense and antisense, respectively, both exhibiting a 0.2% difference from the theoretical MW. This data indicates that the RNA exhibits a monomer state *in vitro* and is the correct species under study. The R_h of each species was also determined using DLS, which was, 72.7 and 71.3 Å for the sense and antisense, respectively [51].

A second example from our research group is the investigation of long-range RNA-RNA interactions in viruses, focusing in on Japanese encephalitis virus (JEV) 5' and 3' terminal regions [18]. Through computational predictions and biophysical validation, the long-range RNA-RNA interacting site of the JEV genome was identified and was also structurally and kinetically characterized. The interaction was measured using several biophysical methods, with SEC-MALS being one of them. Using SEC-MALS, the monomeric RNA MWs were determined, along with the dimerized complex, both determined on SEC-MALS to be 75 kDa for monomers (73.9 and 71.4

kDa theoretical value) and 150 kDa for the complex (145.3 kDa theoretical) (Fig. 2.11) [18]. This result conveyed strong evidence of the species maintaining a monomeric form and not oligomerizing, and further, elucidating that when said monomers interact, they form a dimer.

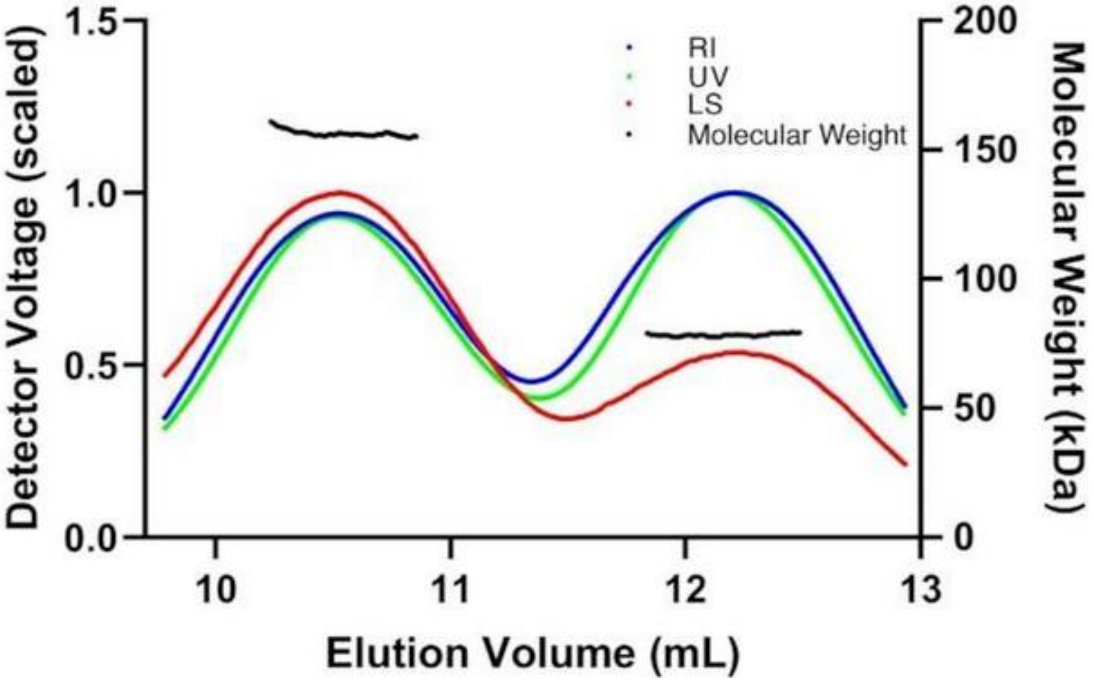


Figure 2.11. SEC-MALS plots of the JEV 5' and 3' terminal regions. The left Gaussian trace is the complex of the RNA species, and the left is the monomers. A 1:1 mixture of RNA was performed and ran at room temperature. The black lines indicate the calculated molecular weights of each separated species [18].

Lastly, a different research group was able to recently characterize the lipid nanoparticle (LNP)-RNA loading occurring using SEC-MALS [101]. The group wanted to establish the effectiveness of assessing the LNP-encapsulated RNA as a quality control and optimization strategy. They used cryogenic electron microscopy (cryo-EM) and DLS and showed synergistic results that found that in non-physiological states in cryo-EM and a semi-physiological state, DLS, they would maintain the same shape and size in solution when retrieving the R_h value [101].

Altogether these examples demonstrate the effectiveness of SEC-MALS/DLS in determining solution properties of a broad range of biomolecules such as MW's, and R_h, as well as the capabilities it possesses in determining this with reliable separation of oligomers and/or interacting species. However, as with every instrument used in scientific discovery, some limitations and considerations must be recognized.

2.17 Considerations

As we consider the limitations of SEC-MALS, it is imperative that when reflecting on your scientific goals SEC-MALS/DLS is primarily a complementary technique to achieve any broad, overarching research objectives. Furthermore, the size exclusion portion of the instrument is limited in separation capability, as with any SEC instrument; any species similar in shape, hydrodynamic properties, and/or size will not be able to be properly separated, resulting in invalid data analysis [102]. If this is a realistic issue for the researcher, then an additional instrument, a field-flow fractionator would be ideal, but this approach could become costly depending on the research groups' monetary affordability [103]. Moreover, due to the high concentration required for some biomolecules based on size and/or scattering potential, the sample can become aggregated [102]. Optimization of buffer conditions is needed to combat this, but for some biomolecules like proteins, it can be difficult to have buffer flexibility to have a folded and/or functional protein passing through the machine for characterization [102]. This also alludes to the fact that MALS/DLS is complementary and requires other techniques to validate the MWs and R_h's of the species under study.

As mentioned earlier, in order to determine the $\left(\frac{dn}{dc}\right)$ value for your species, the current approaches are to look through previous literature, optimize and create a standardized value using the refractive index, or use the batch approach [71, 72]. Collectively, these steps can take time and money, or be out of reach due to previous literature not being available for molecules novel to the SEC-MALS. Due to this, much time can be invested in discovering the $\left(\frac{dn}{dc}\right)$ as again, the accuracy of this value is critical for determining the MW and R_h of the species [71, 72]. Furthermore, many assumptions are made for the R_h value, which describes the size of the particle in solution as a sphere, with no specific hydrodynamics incorporated [71, 72]. These assumptions can work for most molecules, but for some such as the ADH tetramer and kinase fragments do not allow for an accurate R_h determination [104]. Hence, further alluding to SEC-MALS being a complementary technique that can be applied to other complementary data. With all of this in mind, SEC-MALS/DLS offers valuable insights into the solution dynamics of biomolecules by determining oligomerization potential, MWs, and R_h, bringing a strong instrument to use in biophysical characterization. Moreover, keeping research goals in mind its limitations, it warrants the integration of other characterization techniques that altogether bring synergy to the researchers' scientific goals and results.

2.18 References

1. Jung, H., et al., *Ludwig–Soret effect in a linear temperature field: Theory and experiments for steady state distributions.* Physics Letters A, 2011. **375**(19): p. 1917-1920.
2. Ludwig, C., *Difusion awischen ungleich erwärmten Orten gleich zusammengesetzter Lösungen.* Sitz Math Naturwiss Classe Kaiserichen Akad Wiss, 1856. **20**: p. 539.
3. Talbot, L., et al., *Thermophoresis of particles in a heated boundary layer.* Journal of Fluid Mechanics, 1980. **101**(4): p. 737-758.
4. Jerabek-Willemsen, M., et al., *Molecular Interaction Studies Using Microscale Thermophoresis.* ASSAY and Drug Development Technologies, 2011. **9**(4): p. 342-353.
5. Breen, C.J., M. Raverdeau, and H.P. Voorheis, *Development of a quantitative fluorescence-based ligand-binding assay.* Scientific Reports, 2016. **6**(1): p. 25769.
6. Kim, Y.K., et al., *The binding of fluorophores to proteins depends on the cellular environment.* Angew Chem Int Ed Engl, 2011. **50**(12): p. 2761-3.
7. Mrozowich, T., et al., *Investigating RNA–RNA interactions through computational and biophysical analysis.* Nucleic Acids Research, 2023. **51**(9): p. 4588-4601.
8. Zimmermann, R.A., et al., *An Optimized Microscale Thermophoresis Method for High-Throughput Screening of DNA Methyltransferase 2 Ligands.* ACS Pharmacology & Translational Science, 2022. **5**(11): p. 1079-1085.
9. Romain, M., et al., *Measurement of Protein-Protein Interactions through Microscale Thermophoresis (MST).* Bio Protoc, 2020. **10**(7): p. e3574.
10. Pereira, H.S., et al., *Mapping and characterization of G-quadruplexes in monkeypox genomes.* Journal of Medical Virology, 2023. **95**(5): p. e28783.
11. Mueller, A.M., et al., *MicroScale Thermophoresis: A Rapid and Precise Method to Quantify Protein-Nucleic Acid Interactions in Solution.* Methods Mol Biol, 2017. **1654**: p. 151-164.
12. Jerabek-Willemsen, M., et al., *MicroScale Thermophoresis: Interaction analysis and beyond.* Journal of Molecular Structure, 2014. **1077**: p. 101-113.
13. Wienken, C.J., et al., *Protein-binding assays in biological liquids using microscale thermophoresis.* Nature Communications, 2010. **1**(1): p. 100.
14. Nelson, C.R., et al., *Human DDX17 Unwinds Rift Valley Fever Virus Non-Coding RNAs.* Int J Mol Sci, 2020. **22**(1).
15. Nelson, C., et al., *Human DDX3X Unwinds Japanese Encephalitis and Zika Viral 5' Terminal Regions.* Int J Mol Sci, 2021. **22**(1).
16. Chen, M.C., et al., *Insights into the mechanism of a G-quadruplex-unwinding DEAH-box helicase.* Nucleic Acids Res, 2015. **43**(4): p. 2223-31.
17. Hossain, K.A., et al., *Mechanism of recognition of parallel G-quadruplexes by DEAH/RHAU helicase DHX36 explored by molecular dynamics simulations.* Computational and Structural Biotechnology Journal, 2021. **19**: p. 2526-2536.
18. Mrozowich, T., et al., *Investigating RNA-RNA interactions through computational and biophysical analysis.* Nucleic Acids Res, 2023. **51**(9): p. 4588-4601.
19. *A: Scattering and absorption cross-sections*, in *Elements of Modern X-ray Physics*. 2011. p. 343-348.

20. Graewert, M.A., et al. *Adding Size Exclusion Chromatography (SEC) and Light Scattering (LS) Devices to Obtain High-Quality Small Angle X-Ray Scattering (SAXS) Data*. Crystals, 2020. **10**, DOI: 10.3390/cryst10110975.
21. Giannini, C., et al. *X-ray Diffraction: A Powerful Technique for the Multiple-Length-Scale Structural Analysis of Nanomaterials*. Crystals, 2016. **6**, DOI: 10.3390/cryst6080087.
22. Förster, A., S. Brandstetter, and C. Schulze-Briese, *Transforming X-ray detection with hybrid photon counting detectors.* Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2019. **377**(2147): p. 20180241.
23. Kikhney, A.G. and D.I. Svergun, *A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins.* FEBS Letters, 2015. **589**(19, Part A): p. 2570-2577.
24. Forster, S., L. Apostol, and W. Bras, *Scatter: software for the analysis of nano- and mesoscale small-angle scattering.* Journal of Applied Crystallography, 2010. **43**(3): p. 639-646.
25. Hopkins, J.B., R.E. Gillilan, and S. Skou, *BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis.* J Appl Crystallogr, 2017. **50**(Pt 5): p. 1545-1553.
26. Manalastas-Cantos, K., et al., *ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis.* J Appl Crystallogr, 2021. **54**(Pt 1): p. 343-355.
27. Pouget, J.-P., A.-M. Levelut, and J.-F. Sadoc, *André Guinier: Local order in condensed matter.* Comptes Rendus Physique, 2019. **20**(7): p. 725-745.
28. Kratky, O. and G. Porod, *Röntgenuntersuchung gelöster Fadenmoleküle.* Recueil des Travaux Chimiques des Pays-Bas, 1949. **68**(12): p. 1106-1122.
29. Liu, H. and P.H. Zwart, *Determining pair distance distribution function from SAXS data using parametric functionals.* J Struct Biol, 2012. **180**(1): p. 226-34.
30. Choi, K.H. and M. Morais, *Use of small-angle X-ray scattering to investigate the structure and function of dengue virus NS3 and NS5.* Methods Mol Biol, 2014. **1138**: p. 241-52.
31. Franke, D., et al., *ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions.* J Appl Crystallogr, 2017. **50**(Pt 4): p. 1212-1225.
32. Franke, D. and D.I. Svergun, *DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering.* J Appl Crystallogr, 2009. **42**(Pt 2): p. 342-346.
33. Svergun, D.I., *Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing.* Biophys J, 1999. **76**(6): p. 2879-86.
34. Volkov, V.V. and D.I. Svergun, *Uniqueness of ab initio shape determination in small-angle scattering.* Journal of Applied Crystallography, 2003. **36**(3 Part 1): p. 860-864.
35. Panjkovich, A. and D.I. Svergun, *SASpy: a PyMOL plugin for manipulation and refinement of hybrid models against small angle X-ray scattering data.* Bioinformatics, 2016. **32**(13): p. 2062-4.
36. Baronti, L., et al., *A guide to large-scale RNA sample preparation.* Analytical and Bioanalytical Chemistry, 2018. **410**(14): p. 3239-3252.
37. Jackson, R.W., C.M. Smathers, and A.R. Robart, *General Strategies for RNA X-ray Crystallography.* Molecules, 2023. **28**(5): p. 2111.

38. Meier-Stephenson, V., et al., *Identification and characterization of a G-quadruplex structure in the pre-core promoter region of hepatitis B virus covalently closed circular DNA.* J Biol Chem, 2021. **296**: p. 100589.
39. Campbell, O. and P.J. Krause, *The emergence of human Powassan virus infection in North America.* Ticks Tick Borne Dis, 2020. **11**(6): p. 101540.
40. Matthews, V., et al., *Morphological features of Murray Valley encephalitis virus infection in the central nervous system of Swiss mice.* Int J Exp Pathol, 2000. **81**(1): p. 31-40.
41. Kurane, I., *The effect of global warming on infectious diseases.* Osong Public Health Res Perspect, 2010. **1**(1): p. 4-9.
42. Hassett, E.M. and S. Thangamani, *Ecology of Powassan Virus in the United States.* Microorganisms, 2021. **9**(11).
43. Svergun, D., C. Barberato, and M.H.J. Koch, *CRYSOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates.* Journal of Applied Crystallography, 1995. **28**(6): p. 768-773.
44. Kozin, M.B. and D.I. Svergun, *Automated matching of high- and low-resolution structural models.* Journal of Applied Crystallography, 2001. **34**(1): p. 33-41.
45. Mrozowich, T., et al., *Nanoscale Structure Determination of Murray Valley Encephalitis and Powassan Virus Non-Coding RNAs.* Viruses, 2020. **12**(2).
46. Rufini, A., et al., *Senescence and aging: the critical roles of p53.* Oncogene, 2013. **32**(43): p. 5129-43.
47. Amaral, J.D., et al., *The role of p53 in apoptosis.* Discov Med, 2010. **9**(45): p. 145-52.
48. Chillón, I. and A.M. Pyle, *Inverted repeat Alu elements in the human lincRNA-p21 adopt a conserved secondary structure that regulates RNA function.* Nucleic Acids Res, 2016. **44**(19): p. 9462-9471.
49. Xu, Y., et al., *New Insights into the Interplay between Non-Coding RNAs and RNA-Binding Protein HnRNPK in Regulating Cellular Functions.* Cells, 2019. **8**(1).
50. Makeyev, A.V. and S.A. Liebhaber, *The poly(C)-binding proteins: a multiplicity of functions and a search for mechanisms.* Rna, 2002. **8**(3): p. 265-78.
51. D'Souza, M.H., et al., *Biophysical characterisation of human LincRNA-p21 sense and antisense Alu inverted repeats.* Nucleic Acids Res, 2022. **50**(10): p. 5881-5898.
52. Gräwert, M. and D. Svergun, *A beginner's guide to solution small-angle X-ray scattering (SAXS).* The Biochemist, 2020. **42**(1): p. 36-42.
53. Meisburger, S.P., et al., *Breaking the radiation damage limit with Cryo-SAXS.* Biophys J, 2013. **104**(1): p. 227-36.
54. Stachowski, T.R., M.E. Snell, and E.H. Snell, *SAXS studies of X-ray induced disulfide bond damage: Engineering high-resolution insight from a low-resolution technique.* PLoS One, 2020. **15**(11): p. e0239702.
55. Chacón, P., et al., *Low-Resolution Structures of Proteins in Solution Retrieved from X-Ray Scattering with a Genetic Algorithm.* Biophysical Journal, 1998. **74**(6): p. 2760-2775.
56. Grant, T.D., et al., *Small angle X-ray scattering as a complementary tool for high-throughput structural studies.* Biopolymers, 2011. **95**(8): p. 517-30.
57. Andrews, S.S. and J. Tretton, *Physical Principles of Circular Dichroism.* Journal of Chemical Education, 2020. **97**(12): p. 4370-4376.

58. Bishop, G.R. and J.B. Chaires, *Characterization of DNA structures by circular dichroism.* Curr Protoc Nucleic Acid Chem, 2003. **Chapter 7**: p. 7.11.1-7.11.8.
59. Siligardi, G. and R. Hussain, *Applications of Circular Dichroism*, in *Encyclopedia of Spectroscopy and Spectrometry (Second Edition)*, J.C. Lindon, Editor. 2010, Academic Press: Oxford. p. 9-14.
60. Greenfield, N.J., *Using circular dichroism spectra to estimate protein secondary structure.* Nat Protoc, 2006. **1**(6): p. 2876-90.
61. Inc., J. *Circular Dichroism Spectroscopy*. 2023; Available from: https://jascoinc.com/learning-center/theory/spectroscopy/circular-dichroism-spectroscopy/.
62. Wei, Y., A.A. Thyparambil, and R.A. Latour, *Protein helical structure determination using CD spectroscopy for solutions with strong background absorbance from 190 to 230nm.* Biochim Biophys Acta, 2014. **1844**(12): p. 2331-7.
63. Sathyaseelan, C., V. Vijayakumar, and T. Rathinavelan, *CD-NuSS: A Web Server for the Automated Secondary Structural Characterization of the Nucleic Acids from Circular Dichroism Spectra Using Extreme Gradient Boosting Decision-Tree, Neural Network and Kohonen Algorithms.* Journal of Molecular Biology, 2021. **433**(11): p. 166629.
64. Cappannini, A., et al., *NACDDB: Nucleic Acid Circular Dichroism Database.* Nucleic Acids Research, 2023. **51**(D1): p. D226-D231.
65. Micsonai, A., et al., *BeStSel: webserver for secondary structure and fold prediction for protein CD spectroscopy.* Nucleic Acids Research, 2022. **50**(W1): p. W90-W98.
66. Miles, A.J., S.G. Ramalli, and B.A. Wallace, *DichroWeb, a website for calculating protein secondary structure from circular dichroism spectroscopic data.* Protein Science, 2022. **31**(1): p. 37-46.
67. Miles, A.J. and B.A. Wallace, *CDtoolX, a downloadable software package for processing and analyses of circular dichroism spectroscopic data.* Protein Sci, 2018. **27**(9): p. 1717-1722.
68. Micsonai, A., et al., *Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy.* Proc Natl Acad Sci U S A, 2015. **112**(24): p. E3095-103.
69. Ding, J., et al., *Visualizing RNA conformational and architectural heterogeneity in solution.* Nature Communications, 2023. **14**(1): p. 714.
70. Ogawa, T. and N. Hirokawa, *Multiple analyses of protein dynamics in solution.* Biophys Rev, 2018. **10**(2): p. 299-306.
71. Mhatre, R. and I.S. Krull, *Determination of on-line differential refractive index and molecular weight via gradient HPLC interfaced with low-angle laser light scattering, ultraviolet, and refractive index detection.* Anal Chem, 1993. **65**(3): p. 283-6.
72. Striegel, A.M., *Specific refractive index increment (∂n/∂c) of polymers at 660 nm and 690 nm.* Chromatographia, 2017. **80**(6): p. 989-996.
73. Andrews, D.L., *Rayleigh Scattering and Raman Effect, Theory☆*, in *Encyclopedia of Spectroscopy and Spectrometry (Third Edition)*, J.C. Lindon, G.E. Tranter, and D.W. Koppenaal, Editors. 2017, Academic Press: Oxford. p. 924-930.
74. Tyndall, J., *IV. On the blue colour of the sky, the polarization of skylight, and on the polarization of light by cloudy matter generally.* Proceedings of the Royal Society of London, 1869(17): p. 223-233.

75. Strutt, J.W., *LVIII. On the scattering of light by small particles.* The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 1871. **41**(275): p. 447-454.
76. Mie, G., *Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen.* Annalen der physik, 1908. **330**(3): p. 377-445.
77. Debye, P., *Zerstreuung von röntgenstrahlen.* Annalen der Physik, 1915. **351**(6): p. 809-823.
78. Einstein, A., . *AmT. d*. 1910, Phy.
79. Smoluchowski, M.M.-k.T.d., *Opaleszenz von Gasen im kritischen Zustande, sowie einiger verwandter Erscheinungen.* Annalen Physik, 1908. **330**: p. 205-226.
80. Einstein, A., *Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt Annalen der Physik 17*. 1905, Berlin.
81. Zimm, B.H., *Molecular theory of the scattering of light in fluids.* The Journal of Chemical Physics, 1945. **13**(4): p. 141-145.
82. Gans, R., *Asymmetry of gas molecules—an article to determine the molecular form.* Annalen Der Physik, 1921. **65**: p. 97-123.
83. Zimm, B.H., *Apparatus and Methods for Measurement and Interpretation of the Angular Variation of Light Scattering; Preliminary Results on Polystyrene Solutions.* The Journal of Chemical Physics, 2004. **16**(12): p. 1099-1116.
84. Allen, G., *Protein: a comprehensive treatise.* (No Title), 1997.
85. Harding, S. and K. Jumel, *Current protocols in protein science*. 1998, J. Wiley and Sons, New York.
86. Spiechowicz, J., et al. *Diffusion Coefficient of a Brownian Particle in Equilibrium and Nonequilibrium: Einstein Model and Beyond*. Entropy, 2023. **25**, DOI: 10.3390/e25010042.
87. Miller, C.C. and J. Walker, *The Stokes-Einstein law for diffusion in solution.* Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 1997. **106**(740): p. 724-749.
88. Zhou, W., et al., *Ultrafast image-based dynamic light scattering for nanoparticle sizing.* Rev Sci Instrum, 2015. **86**(11): p. 115107.
89. Veziroglu, E.M. and G.I. Mias, *Characterizing Extracellular Vesicles and Their Diverse RNA Contents.* Frontiers in Genetics, 2020. **11**.
90. Du, S., et al., *Extracellular vesicles: a rising star for therapeutics and drug delivery.* Journal of Nanobiotechnology, 2023. **21**(1): p. 231.
91. Cho, H.Y., et al., *Synthesis of Biocompatible PEG-Based Star Polymers with Cationic and Degradable Core for siRNA Delivery.* Biomacromolecules, 2011. **12**(10): p. 3478-3486.
92. Parot, J., et al., *Physical characterization of liposomal drug formulations using multi-detector asymmetrical-flow field flow fractionation.* Journal of Controlled Release, 2020. **320**: p. 495-510.
93. Selvaraj, N., et al., *Detailed Protocol for the Novel and Scalable Viral Vector Upstream Process for AAV Gene Therapy Manufacturing.* Human Gene Therapy, 2021. **32**(15-16): p. 850-861.
94. Zhang, J., R.M. Haas, and A.M. Leone, *Polydispersity Characterization of Lipid Nanoparticles for siRNA Delivery Using Multiple Detection Size-Exclusion Chromatography.* Analytical Chemistry, 2012. **84**(14): p. 6088-6096.

95. Wagner, C., et al. *Biophysical Characterization of Adeno-Associated Virus Vectors Using Ion-Exchange Chromatography Coupled to Light Scattering Detectors*. International Journal of Molecular Sciences, 2022. **23**, DOI: 10.3390/ijms232112715.
96. Amartely, H., et al., *Coupling Multi Angle Light Scattering to Ion Exchange chromatography (IEX-MALS) for protein characterization.* Scientific Reports, 2018. **8**(1): p. 6907.
97. Au - Some, D., et al., *Characterization of Proteins by Size-Exclusion Chromatography Coupled to Multi-Angle Light Scattering (SEC-MALS).* JoVE, 2019(148): p. e59615.
98. Deng, J.Z., et al., *Characterization of High Molecular Weight Pneumococcal Conjugate by SEC-MALS and AF4-MALS.* Polymers (Basel), 2022. **14**(18).
99. Gentiluomo, L., et al., *Coupling Multi-Angle Light Scattering to Reverse-Phase Ultra-High-Pressure Chromatography (RP-UPLC-MALS) for the characterization monoclonal antibodies.* Sci Rep, 2019. **9**(1): p. 14965.
100. Sahin, E. and C.J. Roberts, *Size-Exclusion Chromatography with Multi-angle Light Scattering for Elucidating Protein Aggregation Mechanisms*, in *Therapeutic Proteins: Methods and Protocols*, V. Voynov and J.A. Caravella, Editors. 2012, Humana Press: Totowa, NJ. p. 403-423.
101. Jia, X., et al., *Enabling online determination of the size-dependent RNA content of lipid nanoparticle-based RNA formulations.* J Chromatogr B Analyt Technol Biomed Life Sci, 2021. **1186**: p. 123015.
102. Minton, A.P., *Recent applications of light scattering measurement in the biological and biopharmaceutical sciences.* Anal Biochem, 2016. **501**: p. 4-22.
103. Giordani, S., et al., *Field-Flow Fractionation in Molecular Biology and Biotechnology.* Molecules, 2023. **28**(17).
104. *AN1607: Kinase fragments dimerize without oligomerization domains, shown by SEC-MALS.* Wyatt Technologies.

Preface: ***<u>This chapter was published in the Journal of Biochemistry and Cell Biology in Canadian Science Publishing, Volume 98, Number 1, February 2020, and is in its standard journalistic format.</u>*** **Danielle Gemmill wrote the entire manuscript, with Simmone D'souza, Vanessa Meier-Stephenson and Trushar R. Patel contributing for editing and giving suggestions on the manuscript writing.**

Chapter 3. Current Approaches for RNA labelling to identify RNA-binding proteins

Danielle Gemmill[1], Simmone D'souza[1], Vanessa Meier-Stephenson[1,2], and Trushar R. Patel[1-3]*

[1] Alberta RNA Research and Training Institute & Department of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, AB

[2]Department of Microbiology, Immunology and Infectious Disease, Cumming School of Medicine, University of Calgary, Calgary, AB

[3] Li Ka Shing Institute of Virology and Discovery Lab, University of Alberta, Edmonton, AB

3.1 Abstract

RNA is involved in all domains of life, playing critical roles in a host of gene expression processes, host-defense mechanisms, cell proliferation, and diseases. A critical component in many of these events is the ability for RNA to interact with proteins. Over the past few decades, our understanding of such RNA-protein interactions and their importance has driven the search and development of new techniques for the identification of RNA-binding proteins. In determining which proteins bind to the RNA of interest, it is often useful to use the approach where the RNA molecule is the 'bait' and allow it to capture proteins from a lysate or other relevant solution. Here, we review a collection of methods for modifying RNA in order to capture RNA-binding proteins. These include small-molecule modification, the addition of aptamers, DNA-anchoring, and nucleotide substitution. With each, we provide examples of their application, as well as highlight their advantages and potential challenges.

3.2 Introduction

Over the past several decades, great strides have been made in characterizing the roles of many types of RNA-binding proteins. This is due to the development of newer technologies, and the growing body of researchers that are working towards understanding these interactions. **R**NA-**b**inding **p**roteins (RBPs) form **r**ibo**n**ucleo**p**rotein (RNP) complexes with single-stranded or double-stranded RNA by typically recognizing a specific sequence and/or structure [1-4], as well as through unconventional binding modes as recently reviewed by Hentze et al.[5]. Different domains can be present on a variety of RBPs that allow for the recognition of RNAs, some of these include: **R**NA **R**ecognition **M**otifs (RRM), **K**-**h**omology domains (KH), RGG (Arg-Gly-Gly) boxes, zinc fingers, **d**ouble-**s**tranded **R**NA-**b**inding **d**omains (dsRBD), or DEAD-box (Asp-Glu-Ala-Asp) helicase domains[1-3, 6-8].

Although these RBPs have been known to play important cellular roles by interacting with mRNAs [9-11], recent work has demonstrated the importance of RBPs and their interactions with various types of non-coding RNA such as microRNAs, small nucleolar RNAs, long-non-coding RNA, piRNA and viral RNA [12-23]. Formation of RNP complexes act to regulate essential cellular functions from cell division to cell death, transcription to post-translational gene regulation, and downstream functions [7, 24-26]. It is important to note that the majority of these RNAs is regulated by cellular proteins, and the dysregulation of some of these RNAs have been implicated in various diseases such as cardiovascular, neurological, immune system, and cancer [24, 27-32]. Thus, the identification of RBPs and their mode of interactions could allow researchers to elucidate how fundamental cellular processes are governed.

To investigate the interaction between an RNA of interest and RBPs within a cellular environment, researchers often start off with simpler *in vitro* models, due to the complexity of the cellular environment [33]. Within this approach, T7 RNA polymerase is typically employed to *in-vitro* transcribe linearized DNA to produce the RNA of interest. The RNA produced is then purified using size-exclusion chromatography under the non-denaturing conditions to maintain a native fold of the RNA of interest [34-36]. Such monodispersed *in-vitro* transcribed RNA can then be modified via a number of approaches for use in identification of their binding partners. Here, we review the application of small-molecules (e.g. biotin-labelling), aptamers, nucleotide substitution, and DNA anchoring that can be used to label the RNAs of interest in order to detect their protein binding partners. With each method, we also discuss some of the advantages and disadvantages pertaining to their use.

3.3 Small-molecule modifications

The small-molecule modification is one of the popular RNA-labelling methods, where small molecules can be immobilized on a particular substrate, followed by an application of crude sample (*e.g*. cell lysate) containing a mixture of proteins to perform pull-down assays for identification of specific RBPs. Typical small-molecule tags used are biotin, desthiobiotin or digoxigenin (*Figure 3.1*). Using this approach has the overall advantage of simplicity, cost and appears to have minimal influences on the RNA secondary structure necessary for binding [37, 38].

3.3.1 Biotin labelling

Biotin-labelling of RNA is one of the most widely used non-radioactive method for determining RNA-binding proteins (*Figure 3.1A and 3.2*) [39-42]. The two main types of RNA biotin labelling methods are internal labelling and 5' or 3' end-labelling. Internal labelling relies on the use of a specific biotin labelled **n**ucleotide **t**ri**p**hosphate (NTP) during an *in-vitro* transcription [43]. Internal labelling of the target RNA with biotin has the potential downside of RNA misfolding due to steric hindrance, or may result in synthetic RNA-protein complexes that are not observed intracellularly [44]. Thus, here we focus on the 3' or 5' end labelling methods which are added after *in-vitro* transcription (IVT). In end labelling approach, the 3' or 5' terminus of the RNA is tagged with a single molecule of biotin (*Figures 3.1A and 3.2*) and incubated with cell lysate to allow the RNA to form complexes with cellular proteins [45, 46]. The biotinylated RNA and associated RNA-binding proteins are then pulled down using streptavidin beads to which biotin has a high-affinity for (K_d ~ 6.5 x 10^{-14}) [43, 47, 48]. A similar pull-down method was used employing biotinylated apo-B mRNA to purify several mRNA-associated proteins, including the complementation factor involved in Apobec-1 dependent RNA editing [49].

The biotin-streptavidin interaction is one of the strongest known non-covalent biological interactions and provides an efficient capture of potential RBPs. Moreover, the biotin-streptavidin complex has been demonstrated to sustain a wide range of buffer conditions, including different salt concentrations, heat, pH, and proteolytic conditions, thus offering versatility to the method [50-55]. Typically, mass spectrometry is used downstream to identify proteins that could interact with the biotin-labelled RNA [44]. Once identified using mass spectrometry, selected proteins are expressed and purified to validate their interactions with RNA using methods such as **e**lectrophoretic **m**obility **s**hift **a**ssays (EMSA) and filter binding assays.

There are several benefits to using biotin to label RNA. First, this method is very sensitive, requires less time, is easy-to-use, and is cost-effective when compared to the other methods [39]. Moreover, the 3' or 5' biotin end labelling method can be used for RNA probes from 22-1600 nucleotides long while causing a minimal disturbance in RNA secondary structure [39, 56]. It is also important to note that although this method can be used to label longer RNAs, the biotin labelling incubation times may increase and labelling efficiency decreases as RNA probes increase >450 nt, or have a complex structure [57, 58]. Another shortfall is the very high-affinity interaction between biotin and streptavidin, which requires proteins associated with biotin-labelled RNA to be eluted under harsh denaturing conditions, which affects the re-usability of the streptavidin beads [47]. However, the interaction between streptavidin and biotin can also be broken under non-denaturing conditions by incubating the column in a non-ionic aqueous solution at temperatures above 70ºC for a short period of time [47]. An additional consideration when using this method is that since biotin is a cofactor for many carboxylases in the cell, such as pyruvate carboxylases, there may be false positives with proteins that bind biotin directly [59].

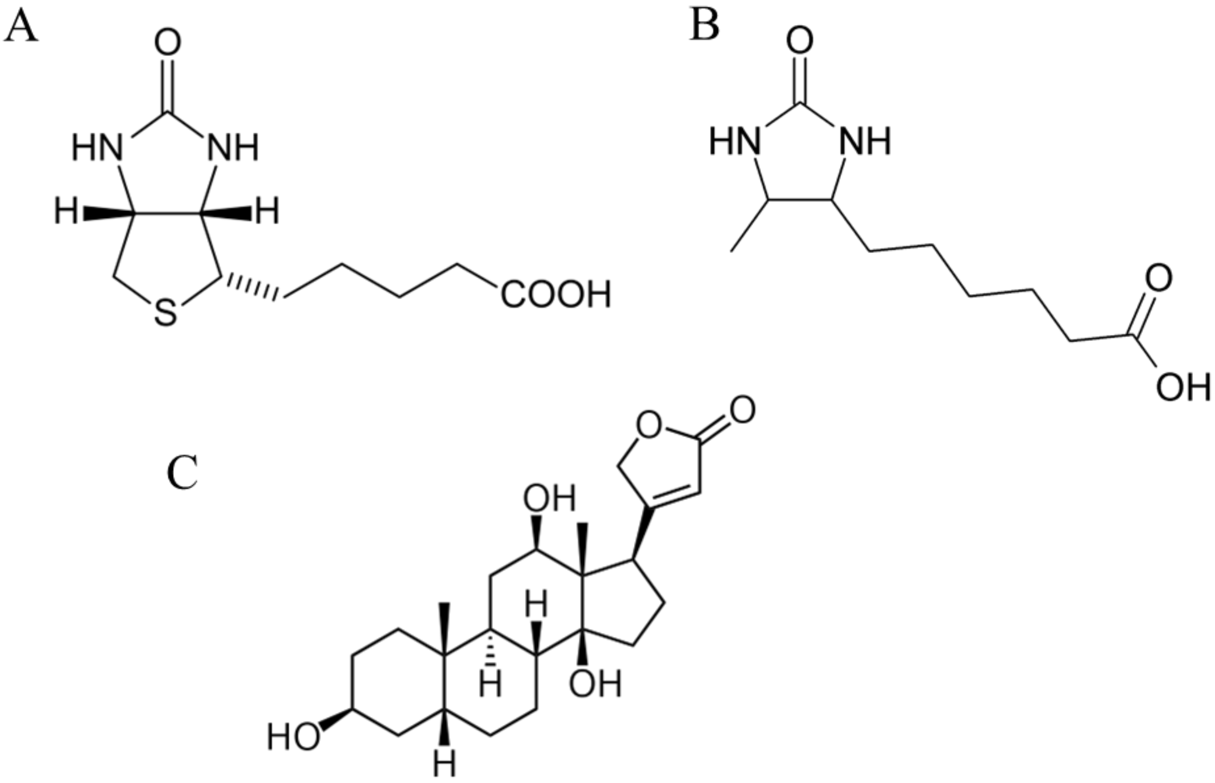


Figure 3.1 Small-molecules typically used to label RNAs to identify RNA-binding proteins. Structure of (A) biotin, (B) desthiobiotin, (C) digoxigenin.

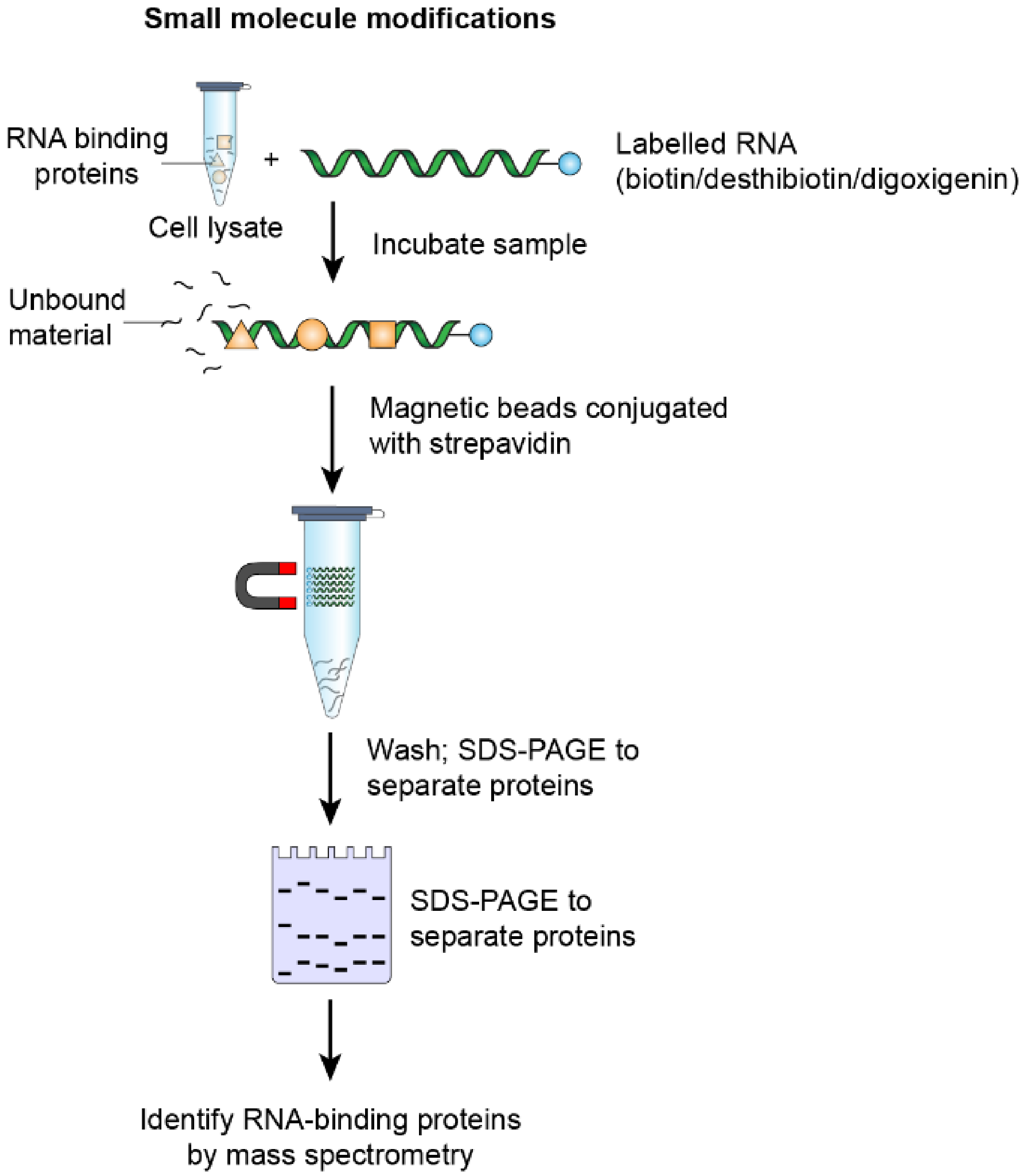


Figure 3.2 An outline of method to capture ribonucleoprotein complexes using small molecules. A flowchart representing the small molecule approach to capture ribonucleoprotein complexes. The labeled RNA is incubated in cell lysate to allow RBPs to interact with the RNA. Biotin and desthiobiotin requires streptavidin-coated magnetic beads, while DIG-labelling protocol employs anti-DIG antibodies conjugated onto magnetic beads. They are added to the cell lysate and RNA mixture, then, the unbound material is washed away while a magnet holds the material of

interest. The components are then separated by SDS-PAGE, and the RNA-binding proteins are identified by mass spectrometry.

3.3.2 Desthiobiotin labelling

Desthiobiotin is a non-sulfur-containing analog of biotin (*Figure 3.1B*) and an alternative to biotin labelling [60]. Desthiobiotin has a lower affinity for biotin capture proteins, such as avidin and streptavidin (K_d 6.6 x 10^{-10}) [61]. Although in principle the pull-down method to identify RNA-binding protein is very similar to that of biotin-labelling, the lower affinity to streptavidin exhibited by desthiobiotin allows for less harsh elution conditions to release the RNA with RNA-bound proteins [44]. For example, the complexes can be eluted using desthiobiotin or a biotin solution to displace the desthiobiotin-bound RNA [60]. Desthiobiotin labelling has been used in a variety of applications, from malignant cell lines to lentiviral vector applications to identify RNA-binding proteins [37, 62].

3.3.3 Digoxigenin labelling

Digoxigenin (DIG) is another popular small-molecule being used for labelling of RNA to identify RNA-binding proteins (Fig. 3.1C). DIG was developed primarily as a way to detect nucleic acids and proteins without the use of radioactivity [38]. There are several ways RNA can be DIG-labelled, such as internal labelling through PCR incorporation of rUTPs, nick translation or direct chemical labelling [38]. However, these methods could possibly interrupt RNA structure, and cause either non-specific binding to RBPs or prevent binding. A solution to this issue is to perform DIG end-labelling at the 3’ terminus of the target RNA. For this reaction to occur, the enzyme terminal transferase catalyzes the addition of a single DIG labelled dideoxy-UTP to the 3’ mRNA end (*Figure 3.2*) [63]. The DIG labelled RNA probe can then be incubated with cellular extract, and then placed to incubate onto magnetic beads. These magnetic beads are conjugated with a chemiluminescent anti-DIG antibody which is used to immunoprecipitate the potential RBPs associated with the DIG-labelled RNA [64]. Herrans and Pallás (2004), demonstrated using DIG-labelled riboprobes that the movement protein interacts with the *Prunis necoritic ringsport* viral ssRNA [65].

The main advantage of DIG over biotin labelling is the minimization of false positive RBPs. Where biotin is found rather ubiquitously in many cell types, digoxigenin is a steroid produced specifically in *Digitalis* plants and does not appear to have confounding interactions with RNA-protein identification in other species [66]. Moreover, DIG hybridization is fast, relatively cheap and sensitive, with less background noise during detection [67].

3.4 Aptamer modifications

In addition to small molecules, the application of RNA aptamers is another popular strategy to modify RNA in order to identify RBPs. An RNA aptamer is a folded oligonucleotide that interacts with a target molecule (*Figure 3.3*). RNA aptamers interact with their respective substrates via electrostatic and hydrophobic interactions along their folded structure (*i.e.*, stem-loops, etc.) enabling anchoring [68-71]. The process of determining aptamers and the substrates they bind to, is known as the **s**ystematic **e**volution of **l**igands by **ex**ponential enrichment (SELEX)[68]. SELEX involves generating a library of oligonucleotides and screening each for binding interactions with a bait molecule (*i.e.*, small molecule, protein, oligonucleotide) [72, 73]. Once interactions are identified, the oligonucleotides are cloned, sequenced, and can then be used to anchor the bait RNA for future capture assays [74]. In its first application, the technique was used to identify a high-affinity RNA that interacts with the bacteriophage T4 DNA polymerase [73]. Since then, aptamers have been identified that have affinities for their ligand that are comparable to antibody-antigen affinities [75-78]. One advantage of using the aptamer approach is that they can be designed and directly incorporated into the RNA of interest during *in vitro* and *in vivo* transcription, without the requirement of additional kits or chemical modifications [74]. The most commonly used RNA aptamers are PP7, S1, D8, Tobramycin, Streptomycin, MS2, Csy4 (H29A), and Mango (*Figures* 3.*3* and 3.*4*) [79-81]. More recently developed aptamers, Csy4, based on the CRISPR-nuclease Csy4 system, and the fluorescent-based quadruplex aptamer, Mango, are also discussed. When considering labelling with an aptamer it is important to understand that aptamers could change the native structure of the RNA which in turn could prevent the binding of RNA-binding proteins [82]. While there are no strict rules on where the aptamer can be placed within the RNA molecule, incorporation at the 5'or 3' ends are the most logical sites to minimize the potential distortions in shape and charge distributions.

3.4.1 PP7 aptamer

The PP7 aptamer (*Figure 3.3A*) was identified using the SELEX and was shown to interacts with the PP7 protein coat of the *Pseudomonas aeruginosa* bacteriophage, with high affinity (K_d ~1nM) [77, 83]. It was demonstrated that the phage protein coat interacts with its own RNA genome to assist in controlling translation [77]. Hogg and Collins used this approach to study the RNA-binding proteins involved in the small nucleolar RNA 7SK, which has been shown to play a role in regulating eukaryotic transcription [84, 85]. This experiment was performed by expressing the PP7 coat protein with an N-terminal fusion of two Protein A domains separated by linker for cleavage by the Tobacco Etch Virus protease. The RNA with the PP7 aptamer is incubated with PP7 coat protein to allow for binding, followed by addition of the cellular lysate and an additional incubation step (*Figure 3.4*). Next, a resin containing anti-PP7 antibodies is used to capture the PP7-bound RNA along with its RBPs. The unbound cell extract is washed away, and the complexes are eluted using TEV protease to cleave the linker [85]. A second purification step can be performed using the tobramycin aptamer by incubating it in agarose derivatized with tobramycin [85]. Elution was then performed using tobramycin in excess. The strength of this approach is that this technique can work with a wide range of ionic strengths and pH [86].

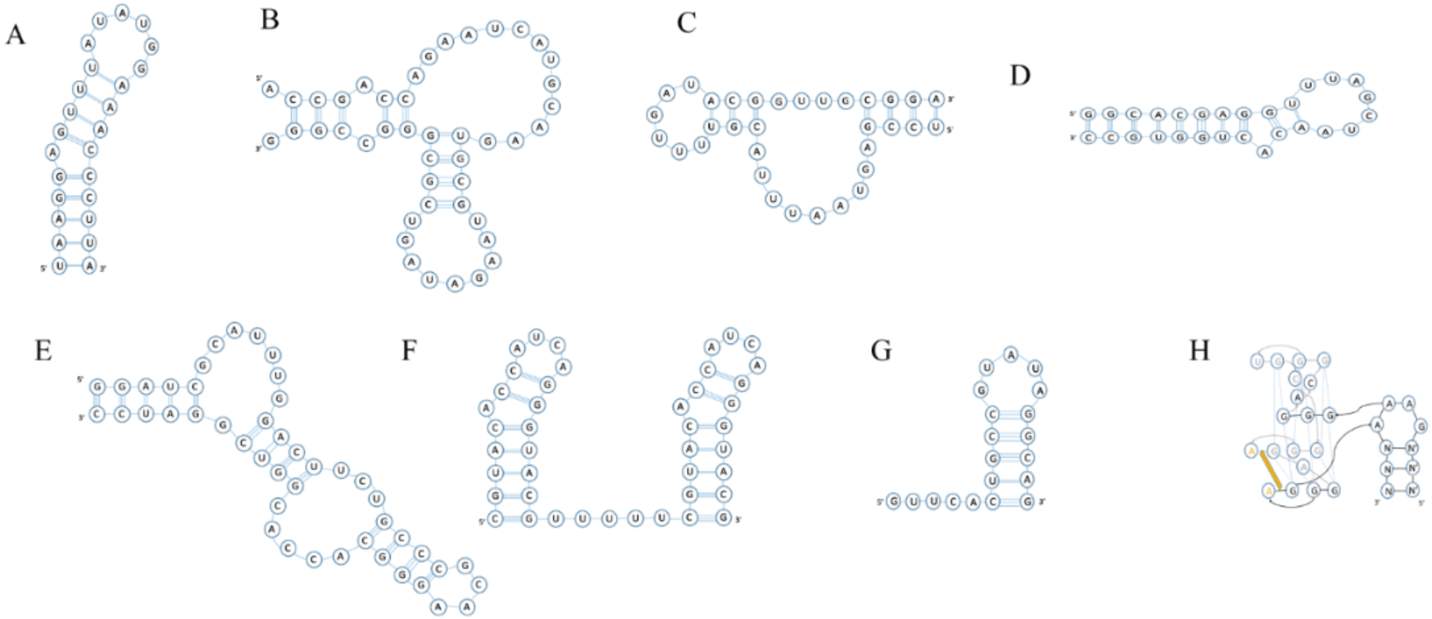


Figure 3.3 Schematics of aptamers that are being used to label RNA molecules. (A) PP7-, (B) S1-, (C) D8-, (D) tobramycin-, (E) streptomycin-, (F) MS2-, (G) Csy4 (H29A)-, and (H) Mango-aptamers.

3.4.2 S1 aptamer

The S1 aptamer (*Figure 3.3B*) has also been identified through SELEX and shown to interact with streptavidin. Dong et al. (2015) performed a pull-down assay with the dengue virus 5’ and 3’ untranslated region (UTR) by modifying the UTRs’ 3’ end with an S1 aptamer, and used streptavidin linked to magnetic beads [87]. The RNA and the proteins bound to it were then eluted using biotin that competed with the streptavidin binding site (*Figure 3.4*). The RNA-binding proteins were then identified using mass spectrometry. S1 was also shown to bind with high affinity to streptavidin when an optimized four tandem repeats are introduced on the RNA for optimal RBP capture [88]. Comparing one S1 aptamer vs. four repeats, Leppek and Stoecklin (2014) observed a 15-fold increase in the captured RNA. However, when using 6 repeats Lioka et. al noticed a loss in efficiency [89]. The four-tandem repeat was shown to bind to streptavidin with a K_d of ~70nM, but with a slight modification of S1, the K_d of the new S1 (S1m) changed to 29nM [88]. It is important to note that while using the S1 aptamer, cellular biotin must be blocked in order to have available streptavidin binding sites to ensure efficient capture of the RNA, therefore, avidin is incubated in cellular extracts prior to putting the RNA with the S1 aptamers into the cell extract. The avidin interacts with cellular biotin, which frees up more of the streptavidin binding sites when performing the pull-down assay [90]. One advantage of using the S1 aptamer over MS2 and PP7 is the readily available streptavidin that can be used for capturing, *i.e.*, recombinant coat proteins are not required to be made or synthesized for experimentation.

3.4.3 D8 aptamer

Similar to the aptamers described above, the D8 RNA sequence (*Figure 3.3C*) was also identified using the SELEX, and was demonstrated to bind to Sephadex® G-100 and G-200, commonly used matrices in gel filtration that employ cross-linked dextran B512 with epichlorohydrin [90-94]. One of the benefits of D8 aptamer is that its interaction with such matrices is highly specific [91, 92]. The potential RBPs attached to the RNA of interest can be eluted with either using a denaturing condition (8M urea) or using a non-denaturing condition by means of application of the competitive binder, dextran B512 [91].

Sephadex® G-100 and G-200 are the cost-effective alternate compared to S1 streptavidin magnetic beads [91]. They are also very stable and can be regenerated and re-used several times [92]. Another benefit to use of this aptamer is that large amounts of cell extract can be used to purify RNA. Additionally, and importantly, the rapid, mild elution conditions make this method very useful for weaker protein-RNA interactions, as well as for preserving native protein and RNA structures during elution [91]. One disadvantage of using the D8 aptamer is its lower affinity for its matrix compared with that of the streptavidin-binding RNA aptamers above [90]. This leads to a balance whereby employing multiple washes results in a loss of D8 bound RNA, but insufficient washes produce more background contamination. The D8 aptamer has been used to identify RNaseP ribozyme from yeast [90, 92] as well as study the protein interactions of the non-coding regions of polioviruses [95].

3.4.4 Tobramycin aptamer

Tobramycin is an antibiotic that interacts with the 30S and 50S ribosomal RNA (rRNA) on the ribosome preventing the 70S complex from forming prior to translation, which consequently causes cell death in bacteria [96]. Using SELEX, the tobramycin-binding aptamer was identified (*Figure 3.3D*) which has a dissociation constant of 5nM with tobramycin [97]. The RNA-labelling assay involves attaching the tobramycin aptamer on the 3' end of adenovirus-derived MINX pre-mRNA. Employing this aptamer, researchers were able to identify a number of proteins involved in RNA splicing, including U1 and U2 snRNP proteins along with non-snRNP splicing factors [98].

3.4.5 Streptomycin aptamer

Similar to tobramycin, streptomycin (*Figure 3.3E*) is also an antibiotic, which binds to the 30S ribosomal subunit of bacteria, interfering with transfer RNA (tRNA) binding and resulting in codon misreads and ultimately cell death [78, 99]. The streptomycin aptamer was found using affinity chromatography with streptomycin-coupled Sepharose^{TM} beads and a pool of nucleotide fragments (*Figure 3.4*) [100]. The K_d of streptomycin aptamer with streptomycin is ~1µM, which is lower compared to the tobramycin dissociation constant. Using this method, spliceosomal U1A protein was purified from yeast cell extract, small nuclear RNA, yeast group II intron-binding proteins, viral RNA binding proteins, and bacterial ncRNA-binding proteins, and the bacteriophage coat protein MS2 [101].

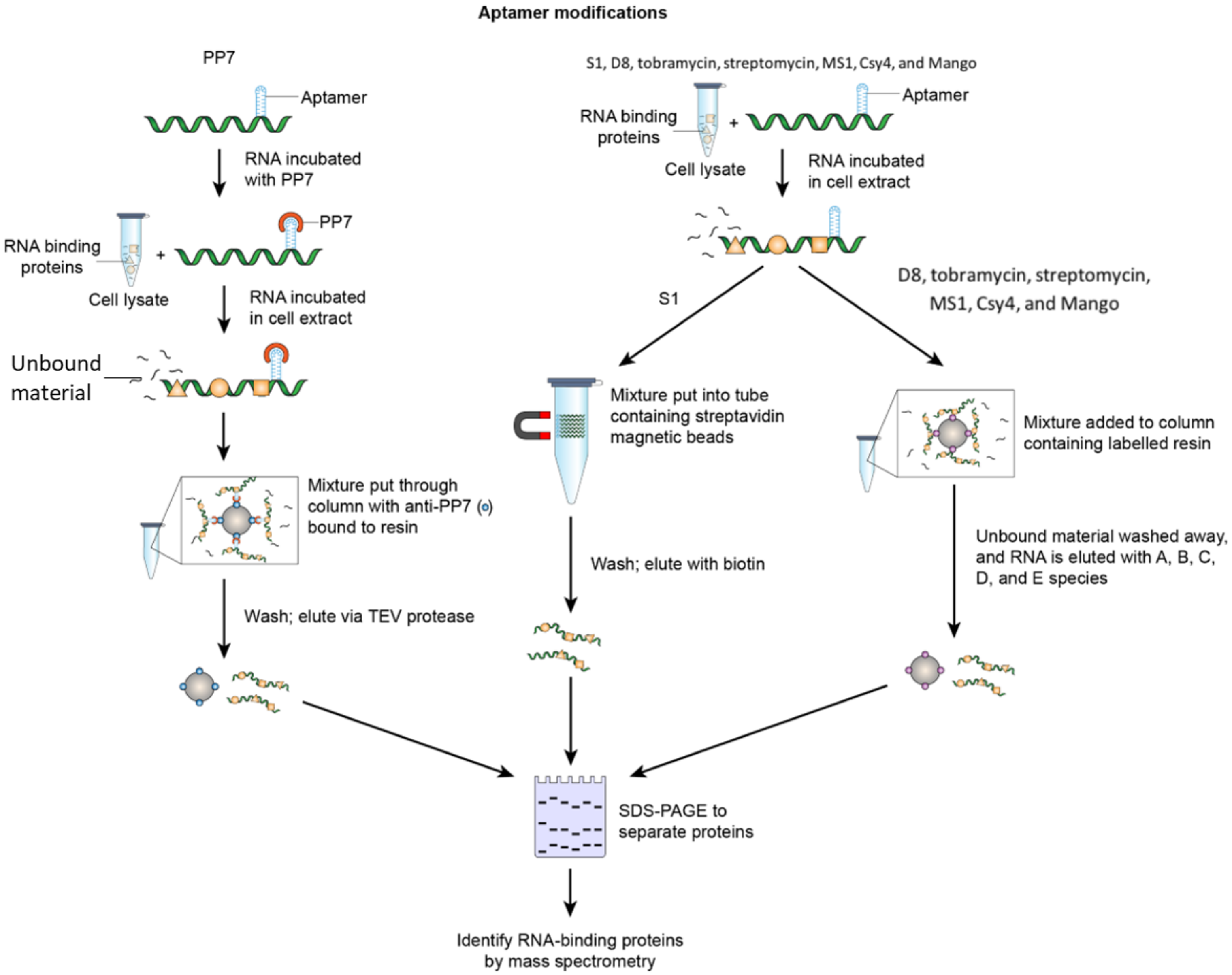


Figure 3.4 Aptamer-based methods for the identification of RBPs. The panel on left presents the PP7 aptamer experiment design using the PP7 coat protein to bind to the aptamer, followed by an incubation with cell lysate. Finally, the sample is incubated in resin conjugated to anti-pp7 antibodies allowing the antibody to bind to the PP7 coat protein. Washing of unbound material is followed by elution with TEV protease cleaving the linker. The middle panel demonstrates the S1 aptamer procedure which uses streptavidin magnetic beads for capturing the ribonucleoprotein complexes. Unbound material is washed away while the magnetic beads hold the streptavidin-ribonucleoprotein complex. Samples are eluted using biotin in excess. Finally, the right figure shows the D8-, tobramycin-, streptomycin-, MS2-, Csy4 (H29A)-, and Mango-aptamers. The resin is labelled with the corresponding capture entity and eluant differs according to the aptamer (8M urea or dextran B512 (D8); tobramycin (tobramycin aptamer); streptomycin

(streptomycin aptamer); MS2 coat protein (MS2); imidazole (Csy4); and biotin (Mango). The D8 aptamer uses Sephadex® G-100 resin to capture the RNP complexes, followed by an elution with 8M urea or dextran B512. Tobramycin and streptomycin aptamers both use their respective antibiotics bound to resin to capture RNP complexes, and elution is performed by adding these antibiotics in excess. Finally, the MS2-aptamer uses the phage MS2 coat protein bound resin to capture the RNP complexes, and elution uses excess MS2 coat protein. All these methods are then carried out on an SDS-PAGE to separate the RNA-binding proteins, and the proteins are identified by mass spectrometry.

3.4.6 MS2 aptamer

The MS2 aptamer is a 19-nucleotide RNA hairpin structure that was derived from the MS2 RNA bacteriophage (*Figure 3.3F*)[76, 102]. The bacteriophage MS2 coat protein interacts with high-specificity and affinity to the MS2 RNA hairpin aptamer [76]. Capitalizing on this strong interaction, multiple tandem repeats (typically runs of 6-8 aptamers) are incorporated into the 3' end of the RNA of interest enabling efficient pulldown using immobilized MS2 coat protein [103, 104]. With either *in-vitro* or *in-vivo* RNA transcription, the MS2 labelled RNA is incubated with cellular extracts or cellular lysate and applied to solid amylose resin (*Figure 3.4*). The resin contains an immobilized fusion protein consisting of the maltose binding protein and the MS2 coat protein [94, 105]. Next, the MS2 labelled RNA and associated RBPs are eluted using excess MS2 coat protein, and the samples can be analyzed using mass spectrometry [83, 106]. The MS2 aptamer has been used to isolate highly stable RNPs, such as U1 snRNP, as well as less stable mRNPs and RNPs associated with non-coding regulatory RNAs [103].

3.4.7 CRISPR endoribonuclease - Csy4

A more recently discovered method of RNA modification is that of tagging with an engineered CRISPR Csy4 endoribonuclease aptamer. Csy4 endoribonuclease is a component of the prokaryotic adaptive immune system, ***c***lustered ***r***egularly ***i***nterspaced ***s***hort ***p***alindromic ***r***epeats (CRISPR), where the nuclease cleaves guide RNA transcripts bearing a specific hairpin sequence (*Figure 3.3G*) [107]. The rationalization for this approach was based on the discovery that a single histidine mutation (H29A) in the enzyme prevents RNA cleavage activity, but that the activity can be readily reversed upon addition of imidazole [107]. Thus, once the Csy4 aptamer-containing RNA is immobilized on the mutant Csy4 endonuclease resin and all unbound cellular components washed away, imidazole can be added, which takes the place of the missing histidine in the active site. This reactivates the enzyme and Csy4 proceeds to cleave the RNA strand, enabling elution of the RNA along with its accompanying RNA-binding proteins. Similar to the above aptamer methods, the subsequent determination of the bound proteins can be undertaken using mass spectrometry [107].

A benefit of this approach is that with the high-affinity binding of the aptamer to the Csy4 nuclease (K_d ~50 pM), more stringent washes can be performed, thereby reducing the background contamination and chances of false positives [82, 107]. The inherent gain in purity via this method removes the need for gel purification before mass spectrometry analysis. Authors claim the mutant Csy4 is a readily produced and robust enzyme, however, would still need to be produced, labelled and incorporated into beads before it could be employed [107]. Other downsides to this approach are that weaker-bound RNA-binding proteins may be washed away by the more stringent washes and therefore lost for identification [107]. This approach has been used to identify microRNA-binding proteins in different cell lines [107], and adaptations of the approach employed in other molecular biology methods [108].

3.4.8 Mango aptamer

The RNA Mango aptamer (*Figure 3.3H*) was designed to provide another option for affinity-based native purification methods of associated RNPs with the added benefit of fluorescence detection [109]. This RNA aptamer consists of a 19nt fluorophore-binding domain that has a high affinity (K_d 3.2nM) to fluorophore reagents: thiazole orange (TO), TO1-biotin or TO1-desthiobiotin [110, 111].The RNA of interest is synthesized by *in vitro* transcription with the Mango aptamer tag and is then incubated in cellular extract to allow for RNP complex formation [109]. The extract containing the Mango-tagged RNA is incubated in streptavidin solid support resin conjugated to TO1-desthiobiotin. Desthiobiotin is used due to its lower binding affinity to streptavidin (K_d 6.6 x 10^{-10}) when compared to biotin (K_d 10^{-14}) [48, 61]. The resin is then washed to remove any non-specific interactions, and the Mango-aptamer-RNA-RNPs are eluted using excess free biotin. The RNP complex can be visualized using SDS-PAGE analysis, further purified using size-exclusion chromatography, and confirm RNA-binding protein identity by LC-MS. Downstream characterization studies between the RNA and RNA-binding protein can also utilize the fluorescent tag by using single-molecule fluorescence cross-correlation spectroscopy, or other methods [109].

The benefits of this method are that the reactions can be performed under native conditions and with similar equipment that is used during DIG labelling (i.e. streptavidin beads, and desthiobiotin). Moreover, the high affinity interaction between Mango and thiazole orange promotes efficient pull down of RNP complexes in nanomolar concentrations with low volumes of cell extract. This approach requires a fluorophore, where the other aptamer methods do not require that additional label. This does give the researcher an added benefit to track the elution of the aptamer complexed with RNA-binding proteins but requires more preparation steps.

3.5 Nucleotide substitution

Another method for RNA anchoring is the application of nucleotide substitution. In this approach, one of the four nucleotides, typically **u**racil **t**riphos**p**hate (UTP) is modified with a heavy metal and incorporated into the RNA during *in vitro* transcription (*Figure 3.5*) [82, 112]. Monoclonal antibodies to this modified uracil can be used to pull down the RNA. The most common nucleotide substitution method for this purpose is the RiboTrap [112].

3.5.1 RiboTrap

RiboTrap is carried out by *in vitro* transcribing the RNA of interest in the presence of 5-bromo-UTP (BrUTP, BrU) free nucleotides [112]. The BrU-labelled RNA is then incubated with cellular lysate. Following incubation, the BrU-RNA-RBP complexes are captured by anti-bromodeoxyuridine (BrdU) monoclonal antibodies which are conjugated to the solid resin [82]. Different washes with varying stringencies can be performed to analyze weak or tightly bound RNA-binding proteins. Elution of the BrU-RNA-bound proteins is with excess BrdU. The pulled-down samples can be run on an SDS-PAGE, and the RNA binding proteins can be identified using mass spectrophotometry [112].

The advantages of this method are that the washes are performed under non-denaturing conditions, and the 5-bromo-UTP RNA modification could minimize non-specific binding [82]. Moreover, this labelling technique can also be used *in vivo* by transfection of the BrU-RNA into yeast hosts, bacteria, and mammalian cell culture to identify RBPs in nuclear and/or cytosolic extracts [112, 113]. The caveat to this modification is that the charge distribution is changed with the substitution of a bromine for the hydrogen in the uracil structure. This, in turn, changes the shape of the molecule and could hinder any native conformations that the RNA possesses. Furthermore, this method requires large amounts of starting material to ensure the detection of RNA-protein interactions [114].

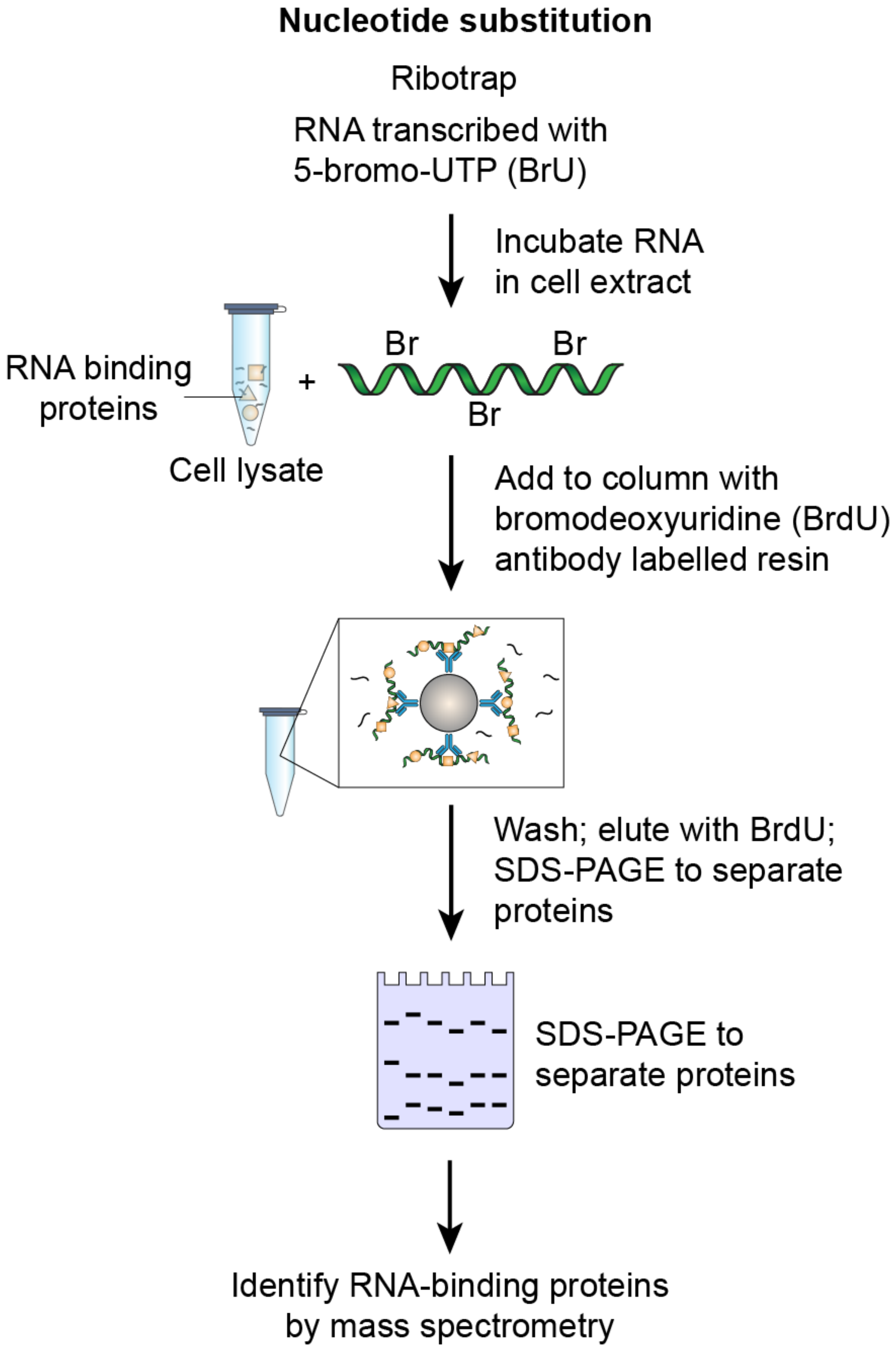


Figure 3.5 Nucleotide substitution approach to identify RBPs. The uracil triphosphates are replaced with 5-bromo-UTP during an *in vitro* transcription process of RNA of interest. The modified RNA is then incubated with cell lysate to allow RBPs to interact with the RNA. The

potential RNA-RBPs complexes are captured by an anti-BrdU antibody that is conjugated to resin. The mixture is loaded onto the column containing appropriate resin, followed by a washing step for removal of unbound components and an elution with excess BrdU to obtain RBPs that can be identified using mass spectrometry.

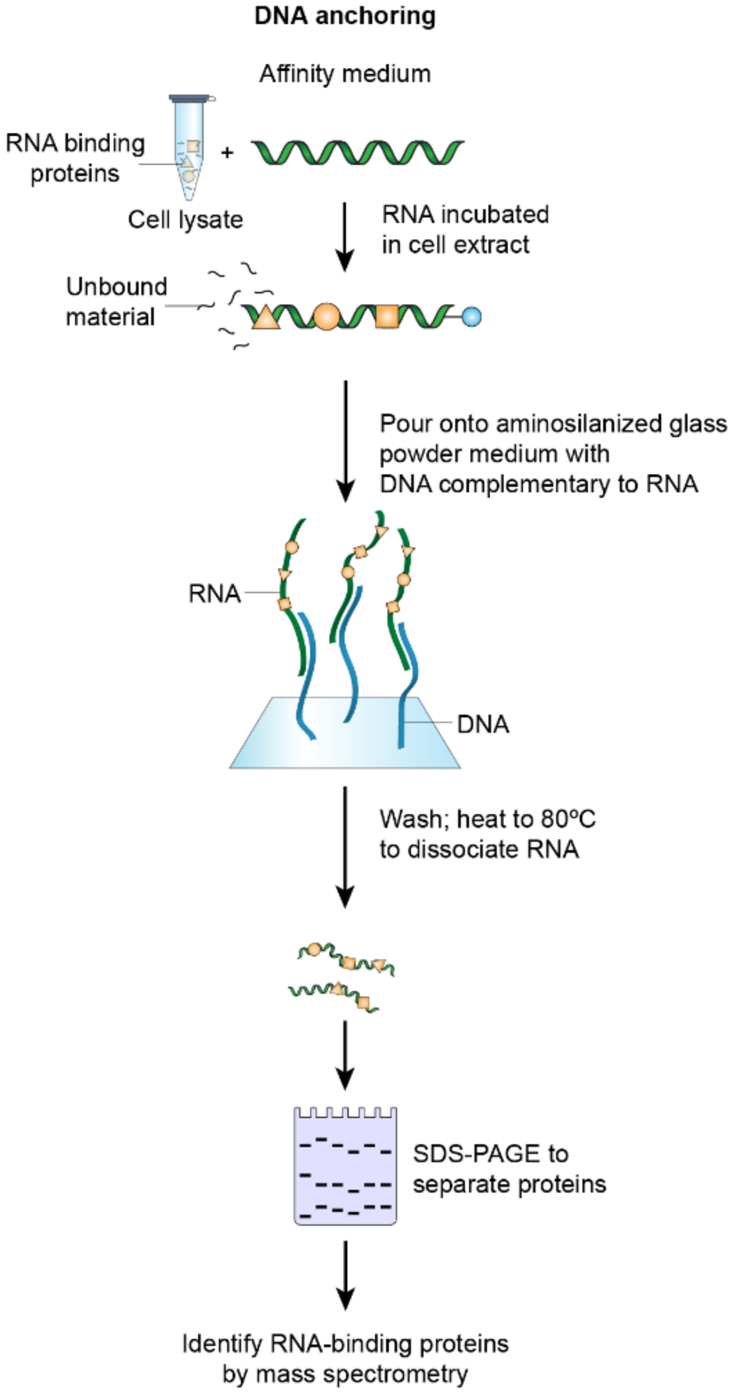


Figure 3.6 The DNA anchoring approach for identifying RBPs. Once the aminosilanized plate is prepared, the ssDNA is ligated onto the plate. The RNA of interest is incubated with cell lysate to allow the formation of ribonucleoprotein complexes. The mixture is then added to the

aminosilanized plate, followed by an incubation step to facilitate interaction between the RNA of interest and ssDNA. The unbound material is removed during a washing step, followed by heating of samples to dissociate the RNA from the ssDNA. As all other methods above, the identification of RBPs is performed using mass spectrometry.

3.6 DNA-anchoring

RNA of interest can also be anchored with the complementary strands of DNA immobilized to a media for identification of RBPs [115, 116]. This approach employs the natural binding affinities of RNA and DNA strands and can incorporate annealing and dissociation temperatures for the capture and elution as described in *Figure 3.6* [115]. An advantage to this approach is that the experiment can be carried out at lower temperatures (4°C) to ensure RNA conformational changes remain in their native state. The disadvantage to DNA anchoring is that RNA-binding proteins that interact weaker to RNA are difficult to capture (Liu and Sun 2005).

3.7 Affinity medium capture

Affinity medium capture approach employs a solid surface to immobilize RNA-binding proteins by creating an aminosilanized glass powder medium or plate, followed by anchoring a strand of ssDNA to that medium with a complementary strand that will base pair to the RNA being studied [116]. The RNA would be modified with a tail that is complementary to the ssDNA anchored to the medium. This allows the RNA to be suspended, allowing RNA-binding proteins to bind to it. The elution is performed by heating the entire complex to 80°C for 5 minutes to denature the DNA-RNA strands. SDS-PAGE and mass spectrometry can be used to identify the proteins and EMSA or other binding studies to confirm RNA binding protein interactions [116].

Three cytoskeletal proteins were identified in this manner revealing *trans*-acting factors that interact with RNA [116].

One of the advantages of this approach is that the whole experiment can be performed at low temperatures, ensuring the RNA conformations stay consistent under physiological or cooler temperatures. As well, the ssDNA glass-powder media can be used to anchor different RNAs that have been extended with the complementary strand, providing a platform for a lab performing multiple of these pulldowns. Additionally, according to the authors, the DNA-glass powder medium can be re-used for multiple RNA pulldowns, implying at least three repeat applications of the same DNA-glass powder plate without compromise of function [116]. The elution step may be disruptive to the native RNA structure, however, and may need to be taken into consideration. One further disadvantage is that weakly bound RNA-binding proteins can be difficult to capture [116].

3.8 Discussion

Major strides have been made in RNA research by using RNA modifications to identify RNA-binding proteins. Further research in this field has improved our understanding of the various biological processes that RNA molecules are involved in. Still, much more work is required in order to understand the dynamics of RNA, and the various regulatory roles employed by RBPs. An application of chemically modified RNA is a powerful tool to answer critically important questions pertaining to RNA's involvement in gene silencing or editing, its involvement in unregulated cell proliferation in cancer, and how viruses use their RNA genome to hijack cellular machinery and propagate themselves [117-119].

Each technique described above has its strengths and challenges. It is important to keep in mind that each approach needs to be optimized for the system under study to identify important RBPs while reducing potential false positive hits. An application of appropriate negative controls (e.g. scrambled nucleotide sequences, or the small molecule or an aptamer by themselves) is critical for the identification of RBPs. Typically, one should also employ more than one technique, where possible, to minimize false positives hits. Such application of multiple techniques will provide common hits identified, which can subsequently be validated using biochemical and biophysical assays in terms of their direct interactions. Furthermore, it is essential to maintain RNA structure during modification to maximize the success of an experiment. As many of these techniques are relatively new discoveries, protocol optimizations and trial and errors may lead to a pricey endeavor. Understanding the benefits and limitations of each of the previously described methods can allow researchers to pick a method that suits experimental needs.

It is important to recognize that there are some inherent challenges associated with studying RNA and RNA-protein interactions. For example, RNase which is present in abundance, is a very robust enzyme that even when autoclaved, it does not completely inactivate [60, 120]. Therefore RNase-free conditions throughout all experiments should be maintained. Next are the challenges associated with the stability of RNA at higher temperatures. Depending on the secondary structure, temperature and duration at elevated temperatures, RNA can be prone to degradation. For example, studies demonstrate that at 25°C, RNA has a degradation rate of 0.7-1.3 nucleotide strand breaks per 1000 nucleotides per century (inferred) when in RNase free containers, while at 90°C had 0.7-7.6 nucleotide strand breaks per 1000 nucleotides per 30 minutes [121]. While heating with subsequent slow cooling is often needed for proper RNA folding *in vitro*, extended periods of time at these higher temperatures can compromise RNA quality and therefore downstream experimentation. Furthermore, when working with RNA, secondary structures and tertiary interactions are typically of utmost importance.

Some of the labelling techniques require temperature increases with the RNA sample, which, as a consequence, could change the native RNA conformation it adopts at physiologic temperatures [82, 121]. As well, the chemical modifications themselves may also influence or hinder the identification of some RNA-binding proteins by mere attachment to a key RNA segment and therefore needs to be taken into consideration when designing experiments.

Overall, there are a multitude of RNA-labelling and anchoring approaches that may be used to identify RBPs. Understanding the advantages and limitations of these methods when interpreting experimental results is of utmost importance. We suggest minimally that a second conformational approach is performed before concluding binding. Given the advances in our knowledge and understanding of the physiological roles of RNA and RNA-binding proteins, the continued drive to study and understand these systems better will be complemented with one or more of the integral techniques described above.

3.9 References

1. Lunde, B.M., C. Moore, and G. Varani, *RNA-binding proteins: modular design for efficient function.* Nat Rev Mol Cell Biol, 2007. **8**(6): p. 479-90.
2. Clery, A., M. Blatter, and F.H. Allain, *RNA recognition motifs: boring? Not quite.* Curr Opin Struct Biol, 2008. **18**(3): p. 290-8.
3. Linder, P. and E. Jankowsky, *From unwinding to clamping - the DEAD box RNA helicase family.* Nat Rev Mol Cell Biol, 2011. **12**(8): p. 505-16.
4. Arnez, J.G. and J. Cavarelli, *Structures of RNA-binding proteins.* Q Rev Biophys, 1997. **30**(3): p. 195-240.
5. Hentze, M.W., et al., *A brave new world of RNA-binding proteins.* Nat Rev Mol Cell Biol, 2018. **19**(5): p. 327-341.
6. Chen, Y. and G. Varani, *Protein families and RNA recognition.* Febs j, 2005. **272**(9): p. 2088-97.
7. Glisovic, T., et al., *RNA-binding proteins and post-transcriptional gene regulation.* FEBS Lett, 2008. **582**(14): p. 1977-86.
8. Valverde, R., L. Edwards, and L. Regan, *Structure and function of KH domains.* Febs j, 2008. **275**(11): p. 2712-26.
9. Dixit, S. and J. Lukes, *Combinatorial interplay of RNA-binding proteins tunes levels of mitochondrial mRNA in trypanosomes.* RNA, 2018. **24**(11): p. 1594-1606.
10. Gerstberger, S., M. Hafner, and T. Tuschl, *A census of human RNA-binding proteins.* Nat Rev Genet, 2014. **15**(12): p. 829-45.
11. Moore, K.S. and M. von Lindern, *RNA Binding Proteins and Regulation of mRNA Translation in Erythropoiesis.* Front Physiol, 2018. **9**: p. 910.
12. Tedeschi, F.A., et al., *The DEAD-box protein Dbp2p is linked to noncoding RNAs, the helicase Sen1p, and R-loops.* RNA, 2018. **24**(12): p. 1693-1705.
13. Lei, W., et al., *Long non-coding RNA SNHG12promotes the proliferation and migration of glioma cells by binding to HuR.* Int J Oncol, 2018. **53**(3): p. 1374-1384.
14. Grzechnik, P., et al., *Nuclear fate of yeast snoRNA is determined by co-transcriptional Rnt1 cleavage.* Nat Commun, 2018. **9**(1): p. 1783.
15. Baou, M., J.D. Norton, and J.J. Murphy, *AU-rich RNA binding proteins in hematopoiesis and leukemogenesis.* Blood, 2011. **118**(22): p. 5732-40.
16. Yang, T.W., et al., *RNA-binding proteomics reveals MATR3 interacting with lncRNA SNHG1 to enhance neuroblastoma progression.* J Proteome Res, 2018.
17. Zhou, F., et al., *HIV-1 Nef-induced lncRNA AK006025 regulates CXCL9/10/11 cluster gene expression in astrocytes through interaction with CBP/P300.* J Neuroinflammation, 2018. **15**(1): p. 303.
18. Qiao, E., et al., *Long noncoding RNA TALNEC2 plays an oncogenic role in breast cancer by binding to EZH2 to target p57(KIP2) and involving in p-p38 MAPK and NF-kappaB pathways.* J Cell Biochem, 2019. **120**(3): p. 3978-3988.
19. Zealy, R.W., et al., *microRNA-binding proteins: specificity and function.* Wiley Interdiscip Rev RNA, 2017. **8**(5).

20. Falaleeva, M., et al., *Dual function of C/D box small nucleolar RNAs in rRNA modification and alternative pre-mRNA splicing.* Proc Natl Acad Sci U S A, 2016. **113**(12): p. E1625-34.
21. Ciafre, S.A. and S. Galardi, *microRNAs and RNA-binding proteins: a complex network of interactions and reciprocal regulations in cancer.* RNA Biol, 2013. **10**(6): p. 935-42.
22. Jiang, P. and H. Coller, *Functional interactions between microRNAs and RNA binding proteins.* Microrna, 2012. **1**(1): p. 70-9.
23. Bagni, C. and B. Lapeyre, *Gar1p binds to the small nucleolar RNAs snR10 and snR30 in vitro through a nontypical RNA binding element.* J Biol Chem, 1998. **273**(18): p. 10868-73.
24. Brinegar, A.E. and T.A. Cooper, *Roles for RNA-binding proteins in development and disease.* Brain Res, 2016. **1647**: p. 1-8.
25. Clayton, C. and M. Shapira, *Post-transcriptional regulation of gene expression in trypanosomes and leishmanias.* Mol Biochem Parasitol, 2007. **156**(2): p. 93-101.
26. Kishore, S., S. Luber, and M. Zavolan, *Deciphering the role of RNA-binding proteins in the post-transcriptional control of gene expression.* Brief Funct Genomics, 2010. **9**(5-6): p. 391-404.
27. Conlon, E.G. and J.L. Manley, *RNA-binding proteins in neurodegeneration: mechanisms in aggregate.* Genes Dev, 2017. **31**(15): p. 1509-1528.
28. Zhou, Q., et al., *The long noncoding RNA LUCAT1 promotes CRC tumorigenesis by targeting RPL40-MDM2-p53 pathway through binding with UBA52.* Cancer Sci, 2019.
29. Newman, A.E.M., et al., *Time as tyrant: The minute, hour and day make a difference for corticosterone concentrations in wild nestlings.* Gen Comp Endocrinol, 2017. **250**: p. 80-84.
30. Li, S., et al., *Long non-coding RNA metastasis-associated lung adenocarcinoma transcript 1 promotes lung adenocarcinoma by directly interacting with specificity protein 1.* Cancer Sci, 2018. **109**(5): p. 1346-1356.
31. Lukong, K.E., et al., *RNA-binding proteins in human genetic disease.* Trends Genet, 2008. **24**(8): p. 416-25.
32. Esteller, M., *Non-coding RNAs in human disease.* Nat Rev Genet, 2011. **12**(12): p. 861-74.
33. Das, D., et al., *Identical RNA-protein interactions in vivo and in vitro and a scheme of folding the newly synthesized proteins by ribosomes.* J Biol Chem, 2012. **287**(44): p. 37508-21.
34. Kanwal, F., et al., *Large-Scale in Vitro Transcription, RNA Purification and Chemical Probing Analysis.* Cell Physiol Biochem, 2018. **48**(5): p. 1915-1927.
35. Easton, L.E., Y. Shibata, and P.J. Lukavsky, *Rapid, nondenaturing RNA purification using weak anion-exchange fast performance liquid chromatography.* RNA, 2010. **16**(3): p. 647-53.
36. Kieft, J.S. and R.T. Batey, *A general method for rapid and nondenaturing purification of RNAs.* RNA, 2004. **10**(6): p. 988-95.
37. Kresoja-Rakic, J. and E. Felley-Bosco, *Desthiobiotin-Streptavidin-Affinity Mediated Purification of RNA-Interacting Proteins in Mesothelioma Cells.* J Vis Exp, 2018(134).
38. McCreery, T., *Digoxigenin labeling.* Mol Biotechnol, 1997. **7**: p. 121-124.
39. Moritz, B. and E. Wahle, *Simple methods for the 3' biotinylation of RNA.* RNA, 2014. **20**(3): p. 421-7.

40. Rinaldi, A.J., K.C. Suddala, and N.G. Walter, *Native purification and labeling of RNA for single molecule fluorescence studies.* Methods Mol Biol, 2015. **1240**: p. 63-95.
41. Liu, Y., et al., *lncRNA GAS5 enhances G1 cell cycle arrest via binding to YBX1 to regulate p21 expression in stomach cancer.* Sci Rep, 2015. **5**: p. 10159.
42. Lindehell, H., M. Kim, and J. Larsson, *Proximity ligation assays of protein and RNA interactions in the male-specific lethal complex on Drosophila melanogaster polytene chromosomes.* Chromosoma, 2015. **124**(3): p. 385-95.
43. Panda, A.C., J.L. Martindale, and M. Gorospe, *Affinity Pulldown of Biotinylated RNA for Detection of Protein-RNA Complexes.* Bio Protoc, 2016. **6**(24).
44. Jazurek, M., et al., *Identifying proteins that bind to specific RNAs - focus on simple repeat expansion diseases.* Nucleic acids research, 2016. **44**(19): p. 9050-9070.
45. Keith, G., *Optimization of conditions for labeling the 3' OH end of tRNA using T4 RNA ligase.* Biochimie, 1983. **65**(6): p. 367-70.
46. England, T.E., A.G. Bruce, and O.C. Uhlenbeck, *Specific labeling of 3' termini of RNA with T4 RNA ligase.* Methods Enzymol, 1980. **65**(1): p. 65-74.
47. Holmberg, A., et al., *The biotin-streptavidin interaction can be reversibly broken using water at elevated temperatures.* Electrophoresis, 2005. **26**(3): p. 501-10.
48. Piran, U. and W.J. Riordan, *Dissociation rate constant of the biotin-streptavidin complex.* J Immunol Methods, 1990. **133**(1): p. 141-3.
49. Mehta, A. and D.M. Driscoll, *A sequence-specific RNA-binding protein complements apobec-1 To edit apolipoprotein B mRNA.* Mol Cell Biol, 1998. **18**(8): p. 4426-32.
50. Gonzalez, M., et al., *Interaction of biotin with streptavidin. Thermostability and conformational changes upon binding.* J Biol Chem, 1997. **272**(17): p. 11288-94.
51. Gonzalez, M., C.E. Argarana, and G.D. Fidelio, *Extremely high thermal stability of streptavidin and avidin upon biotin binding.* Biomol Eng, 1999. **16**(1-4): p. 67-72.
52. Chaiet, L. and F.J. Wolf, *The Properties of Streptavidin, a Biotin-Binding Protein Produced by Streptomycetes.* Arch Biochem Biophys, 1964. **106**: p. 1-5.
53. Hofmann, K., et al., *Iminobiotin affinity columns and their application to retrieval of streptavidin.* Proc Natl Acad Sci U S A, 1980. **77**(8): p. 4666-8.
54. Stayton, P.S., et al., *Streptavidin-biotin binding energetics.* Biomol Eng, 1999. **16**(1-4): p. 39-44.
55. Laitinen, O.H., et al., *Genetically engineered avidins and streptavidins.* Cell Mol Life Sci, 2006. **63**(24): p. 2992-3017.
56. Paingankar, M.S. and V.A. Arankalle, *Identification and characterization of cellular proteins interacting with Hepatitis E virus untranslated regions.* Virus Res, 2015. **208**: p. 98-109.
57. Piskounova, E., et al., *Determinants of microRNA processing inhibition by the developmentally regulated RNA-binding protein Lin28.* J Biol Chem, 2008. **283**(31): p. 21310-4.
58. O'Connor, C.M., C.K. Lai, and K. Collins, *Two purified domains of telomerase reverse transcriptase reconstitute sequence-specific interactions with RNA.* J Biol Chem, 2005. **280**(17): p. 17533-9.

59. Tytgat, H.L.P., et al., *Endogenous biotin-binding proteins: an overlooked factor causing false positives in streptavidin-based protein detection.* Microbial biotechnology, 2015. **8**(1): p. 164-168.
60. Hirsch, J.D., et al., *Easily reversible desthiobiotin binding to streptavidin, avidin, and other biotin-binding proteins: uses for protein labeling, detection, and isolation.* Anal Biochem, 2002. **308**(2): p. 343-57.
61. Levy, M. and A.D. Ellington, *Directed evolution of streptavidin variants using in vitro compartmentalization.* Chem Biol, 2008. **15**(9): p. 979-89.
62. Chen, R., et al., *Affinity recovery of lentivirus by diaminopelargonic acid mediated desthiobiotin labelling.* J Chromatogr B Analyt Technol Biomed Life Sci, 2010. **878**(22): p. 1939-45.
63. Tomassi, A.H., et al., *Nonradioactive Detection of Small RNAs Using Digoxigenin-Labeled Probes.* Methods Mol Biol, 2017. **1640**: p. 199-210.
64. Booy, Evan P., et al., *Comprehensive analysis of the BC200 ribonucleoprotein reveals a reciprocal regulatory function with CSDE1/UNR.* Nucleic Acids Research, 2018. **46**(21): p. 11575-11591.
65. Herranz, M.C. and V. Pallas, *RNA-binding properties and mapping of the RNA-binding domain from the movement protein of Prunus necrotic ringspot virus.* J Gen Virol, 2004. **85**(Pt 3): p. 761-8.
66. McQuaid, S., J. McMahon, and G.M. Allan, *A comparison of digoxigenin and biotin labelled DNA and RNA probes for in situ hybridization.* Biotech Histochem, 1995. **70**(3): p. 147-54.
67. Marchal, A., et al., *A fast, sensitive and cost-effective method for nucleic acid detection using non-radioactive probes.* Biology Methods and Protocols, 2018. **3**(1).
68. Mallikaratchy, P., *Evolution of Complex Target SELEX to Identify Aptamers against Mammalian Cell-Surface Antigens.* Molecules (Basel, Switzerland), 2017. **22**(2): p. 215.
69. Stoltenburg, R., et al., *G-quadruplex aptamer targeting Protein A and its capability to detect Staphylococcus aureus demonstrated by ELONA.* Sci Rep, 2016. **6**: p. 33812.
70. Sun, H., W. Tan, and Y. Zu, *Aptamers: versatile molecular recognition probes for cancer detection.* Analyst, 2016. **141**(2): p. 403-15.
71. Ni, X., et al., *Nucleic acid aptamers: clinical applications and promising new horizons.* Curr Med Chem, 2011. **18**(27): p. 4206-14.
72. Ellington, A.D. and J.W. Szostak, *In vitro selection of RNA molecules that bind specific ligands.* Nature, 1990. **346**(6287): p. 818-22.
73. Tuerk, C. and L. Gold, *Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase.* Science, 1990. **249**(4968): p. 505-10.
74. Keefe, A.D., S. Pai, and A. Ellington, *Aptamers as therapeutics.* Nature Reviews Drug Discovery, 2010. **9**: p. 537.
75. Bjerregaard, N., P.A. Andreasen, and D.M. Dupont, *Expected and unexpected features of protein-binding RNA aptamers.* Wiley Interdiscip Rev RNA, 2016. **7**(6): p. 744-757.
76. Johansson, H.E., L. Liljas, and O.C. Uhlenbeck, *RNA Recognition by the MS2 Phage Coat Protein.* Seminars in Virology, 1997. **8**(3): p. 176-185.
77. Lim, F., T.P. Downey, and D.S. Peabody, *Translational repression and specific RNA binding by the coat protein of the Pseudomonas phage PP7.* J Biol Chem, 2001. **276**(25): p. 22507-13.

78. Demirci, H., et al., *A structural basis for streptomycin-induced misreading of the genetic code.* Nat Commun, 2013. **4**: p. 1355.
79. Bachler, M., R. Schroeder, and U. von Ahsen, *StreptoTag: a novel method for the isolation of RNA-binding proteins.* RNA, 1999. **5**(11): p. 1509-16.
80. Marchese, D., et al., *Advances in the characterization of RNA-binding proteins.* Wiley Interdiscip Rev RNA, 2016. **7**(6): p. 793-810.
81. Ward, A.M., et al., *Quantitative mass spectrometry of DENV-2 RNA-interacting proteins reveals that the DEAD-box RNA helicase DDX6 binds the DB1 and DB2 3' UTR structures.* RNA Biol, 2011. **8**(6): p. 1173-86.
82. Faoro, C. and S.F. Ataide, *Ribonomic approaches to study the RNA-binding proteome.* FEBS Lett, 2014. **588**(20): p. 3649-64.
83. Lim, F. and D.S. Peabody, *RNA recognition site of PP7 coat protein.* Nucleic Acids Res, 2002. **30**(19): p. 4138-44.
84. Peterlin, B.M., J.E. Brogie, and D.H. Price, *7SK snRNA: a noncoding RNA that plays a major role in regulating eukaryotic transcription.* Wiley Interdiscip Rev RNA, 2012. **3**(1): p. 92-103.
85. Hogg, J.R. and K. Collins, *RNA-based affinity purification reveals 7SK RNPs with distinct composition and regulation.* Rna, 2007. **13**(6): p. 868-80.
86. Carey, J. and O.C. Uhlenbeck, *Kinetic and thermodynamic characterization of the R17 coat protein-ribonucleic acid interaction.* Biochemistry, 1983. **22**(11): p. 2610-5.
87. Dong, Y., et al., *Isolation of Endogenously Assembled RNA-Protein Complexes Using Affinity Purification Based on Streptavidin Aptamer S1.* Int J Mol Sci, 2015. **16**(9): p. 22456-72.
88. Leppek, K. and G. Stoecklin, *An optimized streptavidin-binding RNA aptamer for purification of ribonucleoprotein complexes identifies novel ARE-binding proteins.* Nucleic Acids Res, 2014. **42**(2): p. e13.
89. Iioka, H., et al., *Efficient detection of RNA-protein interactions using tethered RNAs.* Nucleic Acids Res, 2011. **39**(8): p. e53.
90. Srisawat, C. and D.R. Engelke, *RNA affinity tags for purification of RNAs and ribonucleoprotein complexes.* Methods, 2002. **26**(2): p. 156-61.
91. Srisawat, C., I.J. Goldstein, and D.R. Engelke, *Sephadex-binding RNA ligands: rapid affinity purification of RNA from complex RNA mixtures.* Nucleic Acids Res, 2001. **29**(2): p. E4.
92. Walker, S.C., et al., *RNA affinity tags for the rapid purification and investigation of RNAs and RNA-protein complexes.* Methods Mol Biol, 2008. **488**: p. 23-40.
93. Ravelet, C. and E. Peyrin, *Aptamers in Affinity Separations: Stationary Separation*, in *Functional Nucleic Acids for Analytical Applications*, L. Yingfu and L. Yi, Editors. 2009, Springer New York: New York, NY. p. 271-286.
94. Slobodin, B. and J.E. Gerst, *A novel mRNA affinity purification technique for the identification of interacting proteins and transcripts in ribonucleoprotein complexes.* RNA, 2010. **16**(11): p. 2277-90.
95. Flather, D., et al., *Generation of Recombinant Polioviruses Harboring RNA Affinity Tags in the 5′ and 3′ Noncoding Regions of Genomic RNAs.* Viruses, 2016. **8**(2): p. 39.

96. Yang, G., et al., *Binding of aminoglycosidic antibiotics to the oligonucleotide A-site model and 30S ribosomal subunit: Poisson-Boltzmann model, thermal denaturation, and fluorescence studies.* J Med Chem, 2006. **49**(18): p. 5478-90.
97. Jiang, L. and D.J. Patel, *Solution structure of the tobramycin–RNA aptamer complex.* Nature Structural Biology, 1998. **5**: p. 769.
98. Hartmuth, K., et al., *Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method.* Proc Natl Acad Sci U S A, 2002. **99**(26): p. 16719-24.
99. Sharma, D., et al., *Mutational Analysis of S12 Protein and Implications for the Accuracy of Decoding by the Ribosome.* Journal of Molecular Biology, 2007. **374**(4): p. 1065-1076.
100. Wallace, S.T. and R. Schroeder, *In vitro selection and characterization of streptomycin-binding RNAs: recognition discrimination between antibiotics.* Rna, 1998. **4**(1): p. 112-23.
101. Windbichler, N. and R. Schroeder, *Isolation of specific RNA-binding proteins using the streptomycin-binding RNA aptamer.* Nat Protoc, 2006. **1**(2): p. 637-40.
102. Carrier, M.C., D. Lalaouna, and E. Masse, *A game of tag: MAPS catches up on RNA interactomes.* RNA Biol, 2016. **13**(5): p. 473-6.
103. Lykke-Andersen, J., M.D. Shu, and J.A. Steitz, *Human Upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon.* Cell, 2000. **103**(7): p. 1121-31.
104. Yoon, J.H., S. Srikantan, and M. Gorospe, *MS2-TRAP (MS2-tagged RNA affinity purification): tagging RNA to identify associated miRNAs.* Methods, 2012. **58**(2): p. 81-7.
105. Zhou, Z. and R. Reed, *Purification of functional RNA-protein complexes using MS2-MBP.* Curr Protoc Mol Biol, 2003. **Chapter 27**: p. Unit 27 3.
106. Said, N., et al., *In vivo expression and purification of aptamer-tagged small RNA regulators.* Nucleic Acids Res, 2009. **37**(20): p. e133.
107. Lee, H.Y., et al., *RNA-protein analysis using a conditional CRISPR nuclease.* Proc Natl Acad Sci U S A, 2013. **110**(14): p. 5416-21.
108. Liang, Y., et al., *Endoribonuclease-Based Two-Component Repressor Systems for Tight Gene Expression Control in Plants.* ACS Synthetic Biology, 2017. **6**(5): p. 806-816.
109. Panchapakesan, S.S.S., et al., *Ribonucleoprotein purification and characterization using RNA Mango.* RNA (New York, N.Y.), 2017. **23**(10): p. 1592-1599.
110. Xia, Y., et al., *Recent advances in high-performance fluorescent and bioluminescent RNA imaging probes.* Chemical Society reviews, 2017. **46**(10): p. 2824-2843.
111. Trachman, R.J., 3rd, et al., *Structural basis for high-affinity fluorophore binding and activation by RNA Mango.* Nature chemical biology, 2017. **13**(7): p. 807-813.
112. Beach, D.L. and J.D. Keene, *Ribotrap : targeted purification of RNA-specific RNPs from cell lysates through immunoaffinity precipitation to identify regulatory proteins and RNAs.* Methods Mol Biol, 2008. **419**: p. 69-91.
113. Han, W., et al., *Ribotrap analysis of proteins associated with FHL3 3'untranslated region in glioma cells.* Chin Med Sci J, 2014. **29**(2): p. 78-84.
114. Michlewski, G. and J.F. Caceres, *RNase-assisted RNA chromatography.* Rna, 2010. **16**(8): p. 1673-8.
115. Au - Torres, M., et al., *RNA Pull-down Procedure to Identify RNA Targets of a Long Non-coding RNA.* JoVE, 2018(134): p. e57379.

116. Liu, D.G. and L. Sun, *Direct isolation of specific RNA-interacting proteins using a novel affinity medium.* Nucleic Acids Res, 2005. **33**(15): p. e132.
117. Hamilton, A., et al., *Two classes of short interfering RNA in RNA silencing.* Embo j, 2002. **21**(17): p. 4671-9.
118. Hyodo, K. and T. Okuno, *Host factors used by positive-strand RNA plant viruses for genome replication.* Journal of General Plant Pathology, 2014. **80**(2): p. 123-135.
119. Jansson, M.D. and A.H. Lund, *MicroRNA and cancer.* Mol Oncol, 2012. **6**(6): p. 590-610.
120. Miyamoto, T., S. Okano, and N. Kasai, *Irreversible thermoinactivation of ribonuclease-A by soft-hydrothermal processing.* Biotechnol Prog, 2009. **25**(6): p. 1678-85.
121. Fabre, A.L., et al., *An efficient method for long-term room temperature storage of RNA.* Eur J Hum Genet, 2014. **22**(3): p. 379-85.

Chapter 4. Identification of potential binding partners interacting with the Zika virus terminal regions - *__Contribution percentage = 100%__*

4.1 Introduction

Viruses have amassed many strategies to proliferate themselves through time. They do so primarily by mutating their genomes to avoid any counteroffensive immunity from their hosts yet maintain the delicate balance of maintaining conserved regions that are critical for propagation/transmission within the hosts they infect [1-4]. The amount of conservation within their genome depends on the type of virus (DNA or RNA), as well as any selection pressures enacted by the host they replicate in [2-4]. Examples of protein structure conservation can be seen in the Influenza A virus, wherein the 8-segmented single-stranded RNA (ssRNA) virus possesses 3 highly conserved segments - the NP, involved in RNA-RNA interactions and viral packaging, the M segment, playing essential roles in viral packaging, and the NS segment, which consist of the NS1 and NS2 genes, which are involved in regulating viral replication, as well as host adaptation and immune evasion [1-7]. Another example can be seen in the RNA-dependent RNA polymerase (RdRp) that many RNA viruses possess (+/- sense, and double-stranded RNA (dsRNA)), such as Flaviviruses, Picornaviruses, coronaviruses, influenza, measles, rabies, reoviruses, and rotaviruses, to name a few [8-15]. There are many motifs and residues that are conserved across each of these families [16-20]. Altogether, this pattern of conservation is all related to optimal protein function for the virus to successfully propagate and transmit itself. Moreover, there are also many examples of regions of viruses that are not necessarily involved in viral protein function but are involved in optimal packaging, as seen in the Foot and Mouth Disease virus (FMDV) [21]. This study showed that specific stem-loop structures in the 3’ untranslated region (UTR) were critical for viral packaging, and any disruption of these loops caused a significant decrease in the assembly of mature virions [21]. It was also identified in the Influenza A virus that the packaging stem-loop 2 (PSL2) is directly involved in viral packaging, and when the researchers targeted this region with mice infected with Influenza A with a locked nucleic acid, 100% of the mice survived, and viral titers decreased as strong immunity was able to develop [22].

4.2 Current global Flavivirus concern

The *Flaviviridae* family are a taxonomic group of +ssRNA viruses that can be further subdivided into what specific host they infect, as this family infects many oceanic and terrestrial species. The mosquito-borne flaviviruses (MBFV) are a cluster of neuroinvasive arboviruses that are the most prevalent type of flavivirus that burden human health globally, and cause weakening economic outcomes with the infection of livestock. Much research has been invested in the development of a vaccine for flaviviruses that infect humans and livestock, and there are currently vaccines available for West Nile for horses, Yellow Fever, Japanese Encephalitis, Tick-Borne Encephalitis, and until very recently, Dengue virus (DENV) for humans. Unfortunately, even though a vaccine may be available, due to economic inequalities/structures, lack of infrastructure, clinical application restriction, or international collective efforts for vaccination distribution, billions of individuals are at risk of contracting flaviviruses and consequently causing permanent neurological sequelae or death. For example, the newly approved DENV live vaccine works against all 4 serotypes of DENV, however, only individuals aged 6-16 qualify for vaccine administration [23]. To secure a future wherein flaviviruses do not impact public health or economic prosperity, multiple facets need to contribute to decrease the aforementioned effects. This includes, but is not limited to, political policy and economic contribution, public health education, and scientific discovery. In general, more research is required to create new tools for targeting these viruses, and more importantly, if there is an antiviral or vaccine that is designed to target multiple MBFV-conserved structures.

4.3 Zika virus public health threat

Over the last decade, MBFVs have increasingly become a public health issue, especially with MBFVs wherein there is no vaccine or treatment for it [24, 25]. Zika virus (ZIKV) has become one of the more prevalent MBFVs, most notably during the 2015/2016 epidemic of ZIKV in South America, when it was discovered that ZIKV cases were not only significantly on the rise but were vertically transmitting in pregnant individuals to the developing fetus, causing gestational microencephaly [24, 25]. According to the World Health Organization, since this significant 2015 epidemic occurred, the virus has been further investigated and shown to be a risk in 89 countries [26]. Based on the harmful neurological sequelae of ZIKV, along with the risk of gestational microencephaly, it remains critical that researchers develop either a vaccine or antiviral to prevent future outbreaks and strain on public health.

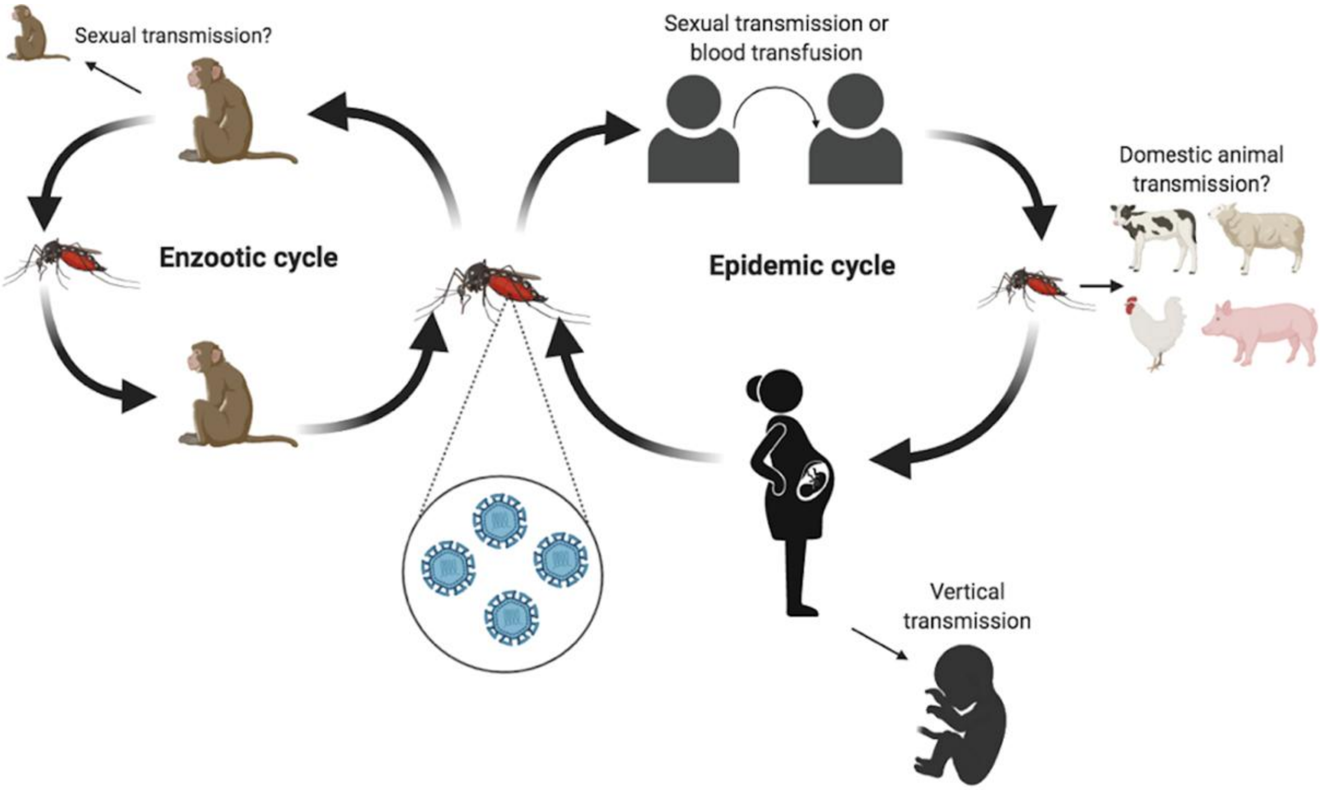


Figure 4.1 The transmission cycles of ZIKV. The virus cycles in the enzootic cycle, maneuvering through vertebrates and invertebrates, allowing amplification of virions. With the enzootic cycle along with the rapid urbanization by *Homo sapiens*, and anthropogenic climate change, mosquito-borne flavivirus amplification and infection (to date) is the highest it's been [27-29]. From the enzootic cycle, the virus can spill over into humans, amplify in a population either through sexual transmission, blood transfusions, gestational vertical transmission, or mosquito-human contact transmitting it back and forth, and completing the amplification cycle. Furthermore, it is unknown if the virus can be transmitted to livestock, as seen in a host of other flaviviruses, such as West Nile virus (created with BioRender.com) [30].

Currently, the antiviral drugs used in targeting focus on viral attachment/entry, uncoating, polymerase, protease, and integrase inhibitors [31]. Additionally, there are also nucleotide analogs, which cause an overall decline in viral replication, alterations in viral exonuclease function, and disturbed proofreading activity [31]. As mentioned above, there are no treatments or vaccines for ZIKV. regarding the absence of a vaccine, it is mainly due to it potentially causing Guillain-Barré Syndrome (GBS) [31-36]. More research is required to find alternative ways to target, and a relatively newer avenue of antiviral research that our research group focuses on is investigating conserved regions of viruses and the host molecular machinery that are hijacked by them to drive viral replication; more specifically, a host proteomics approach of understanding what host proteins are interacting with the virus to help drive replication forward [37-42].

4.4 Conservation of the terminal regions in ZIKV and exploited human proteins

As mentioned above, RNA viruses tend to have less fidelity in their polymerases resulting in a substantial increase in mutations when compared to DNA viruses [43-46]. It is worth noting that viral fitness is crucial as well, as any mutations that do not produce fitness for that moment in time for that host will result in the cessation of viral copies and the risk of extinction [43, 46, 47]. ZIKV has many conserved regions within the genome, with some hypothesized to be involved in optimal protein functionality, RNA folding or packaging, immune evasion as seen with dinucleotide bias with CpG islands, and the 5’ and 3’ non-coding terminal regions (TRs) of the virus [48, 49]. Various sequences and secondary structures of the TRs are highly conserved across all flaviviruses, however, it is unknown what all the various functions of these regions are and what host proteins are hijacked by them [48, 49]. Currently, data has shown that the terminal regions act as molecular switches and are the sole elements responsible for the regulation of transcription and translation of the viral genome [48, 49]. Due to the virus being a +ssRNA, the polymerase and the ribosome cannot bind and simultaneously transcribe and translate respectively, therefore, a molecular switch a required. The TRs in flaviviruses have been shown to provide many functions that not only provide pathogenic enhancement, but also, are critical for the virus to replicate in general and pass from host to host [48-53]. The TRs are critical for the passage between mammals and arthropods and provide the means of host adaptation [48-53]. They also contain XRN-1 5'→3' exoribonuclease resistant regions that allow for the generation of long non-coding viral RNA within the 3' TR called, subgenomic flaviviral RNA (sfRNA), which have been shown to provide a 'protein sponging' function that hijacks multiple proteins, causing the kinetic landscape of the cell to change and favour viral replication more in human cells [54]. In general, it is not fully known what human proteins are being hijacked by sfRNA, or the terminal regions in general. So far it has been identified that splicing and RNA-induced silencing complex (RISC) proteins are hijacked, and as well, as Musashi-1, Fragile X mental retardation protein, Y-box-binding protein 1, and stress granule proteins, but a comprehensive analysis of all the proteins has yet to be discovered, which altogether could provide new insights on how to target ZIKV [52, 55-58].

With the lack of elucidation of all the human proteins hijacked by the 5' and 3' TRs, here, we aimed to identify all proteins that interact with the TRs, directly and indirectly. We did so by performing an immunoprecipitation pulldown assay using the 5' and 3' TRs as bait. Following the pulldowns, mass spectrometry was performed to identify the proteins that were in the final eluting sample. We then selected 5 proteins to validate direct interactions with the viral TRs by: analyzing identified proteins from literature that have a history of interacting with other flaviviruses, the roles the proteins are predicted to play, repetitive and high confidence mass spectrometry results, their size, and ease of expression and purification based on the previous literature. From here we aimed to characterize these interactions using kinetic and biophysical techniques (see Chapters 5 & 6). We will further determine if the characterized proteins are critical for viral replication by performing *in vivo* human cell studies using a ZIKV replicon system.

With the collective information that the TRs are highly conserved, and are critical for replication and host-switching but lacking in identification of human proteins hijacked by them, we hope that this data will provide not only insight into other functions the TRs possess, but also, allow for researchers to utilize them as a treatment target to alleviate the harsh fallout that this virus causes.

4.5 Materials and Methods

4.5.1 *in vitro* transcription, purification, and labelling of ZIKV 5’ and 3’ TRs

Using the TRs as bait in the pulldown assay, we implemented an *in vitro* transcription (IVT) approach by utilizing commercial approaches with Integrated DNA Technologies™. A sequence containing our transcript of interest was cloned into pUC57-Kan vectors containing a T7 promotor in preparation for our in-house IVTs. The vector also contained an *Xba*I cut site to allow the T7 polymerase to dissociate from the DNA at the end of our transcript sequence. The IVTs were performed by mixing transcription buffer (200 mM Tris-Cl (pH 7.5), 75 mM $MgCl_2$, 10 mM spermadine, and 50 mM NaCl), 100 mM DTT, 25 mM NTPs, 100 mM GMP, 0.5U/µL IPPase, 10 µM T7 polymerase, Ribolock™, and the digested plasmid (~600ug/mL added), and was incubated altogether for 3 hours at 37°C. This was followed immediately by purification via size-exclusion chromatography (SEC) using a Cytiva™ Superdex™ 200 Increase 10/300 GL for the 5’ TR and a Sephacryl™ S-400 HR for the 3' TR. The elution fractions were run on a 2% agarose gel. Monomeric fractions containing the appropriate size were collected and used for the digoxigenin (DIG) labelling reaction. Furthermore, this process was also repeated with a designed RNA that was a scrambled sequence form of the viral RNA [59]. This was performed to show both general RNA binding protein partners as well as specific proteins that interact with the ZIKV TRs.

All purified RNA was then subjected to a 3’ labelling reaction of a 5-mer oligonucleotide modified on the 3’ end with the hydroxy steroid molecule, digoxigenin (DIG). This molecule was selected as the immunoprecipitation bait based on two considerations: 1. Due to its high affinity for the DIG antibody (~12 nM), and secondly, that it is only found in leaves and plants in the plant genus, *Digitalis*, to minimize proteins that might bind to DIG in human cells as shown in biotin immunoprecipitation pulldown assays [60, 61]. Altogether, this provided specificity in binding partners, as well as a robust affinity that provides adequate capturing. The 5-mer oligo-DIG species was mixed with 30 µM of 5’ & 3’ TR RNA and incubated in a T4 RNA ligase reaction cocktail that was used by New England Biolabs™ that labeled the 3’ end of the 5’ and 3’ TR transcripts. The reaction was incubated at 12°C for 12 hours, followed by an RNA cleanup that was performed using the Biobasic™ RNA cleanup kit. The purified, modified RNA was then stored for checking via electrophoretic mobility shift assay (EMSA) if the RNA was labelled, as well as for the pulldown assay. The labelling efficiency was checked via 2% agarose gel EMSA with all RNA loaded at a concentration of 1 µM, and anti-DIG IgG antibodies loaded at 10 µM. The shifting in each lane is indicative of an interaction occurring, and the labelling efficiency based on the shifting was determined using ImageJ (Fig. 4.2) [62].

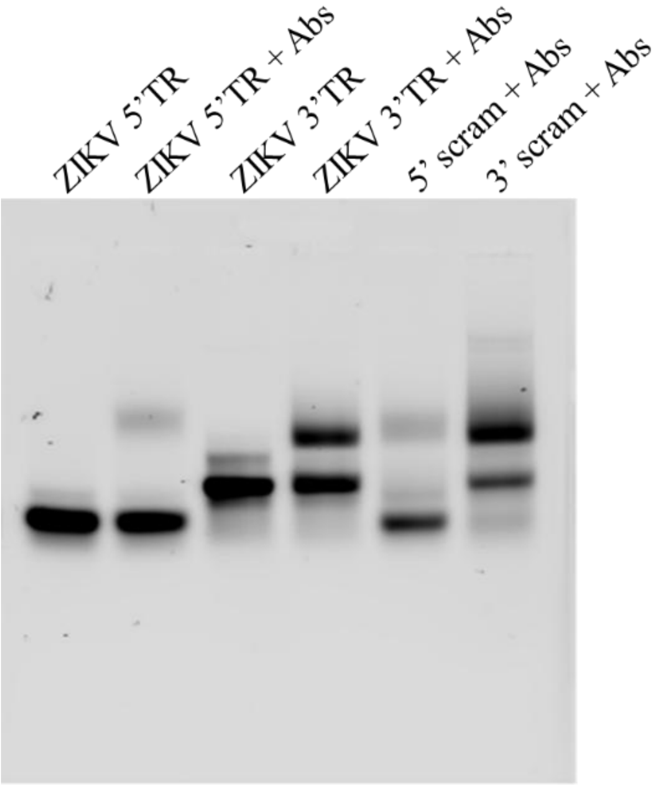


Figure 4.2 A 2% agarose electrophoretic mobility shift assay of 1 µM DIG-labelled 5' and 3' TRs and 5' and 3' Scrambled RNA with 10 µM of anti-digoxin antibodies that were used in the pulldown assays for capturing. The respective labelling efficiencies of the pulldowns using ImageJ were: 31.8% for the 5' TR, 50% for 5' Scrambled, 71.7% for 3' TR, and 87.5% for 3' Scrambled. SYBR™ Safe dye (ThermoFisher Scientific™) was used to stain the nucleic acids for imaging.

4.5.2 Preparation of the SH-SY5Y human cell line

Due to ZIKV neuroinvasivity, the pulldown assay requires neurological relevance to understanding the neuronal RNA-binding protein landscape that binds to the TRs. Thus, the ideal cell line that we used was SH-SY5Y cells. These cells are used in any research that is neuronal focused, such as Parkinson's, Alzheimer's, other neurodegenerative disorders, or any infectious disease that infects neuronal tissues [63-69]. The cells were revived from liquid nitrogen and were grown on 15 cm plates (one plate per pulldown) at 37°C in a media cocktail of F12, Eagle minimal essential media, fetal bovine serum, L-glutamine, non-essential amino acids, and pen/strep. The cells were split using TrypsinX and PBS, followed by replenishing with a media cocktail. The cells were grown until 70-80% confluency, then harvested into a 1.5 mL tube in preparation for the pulldown assay.

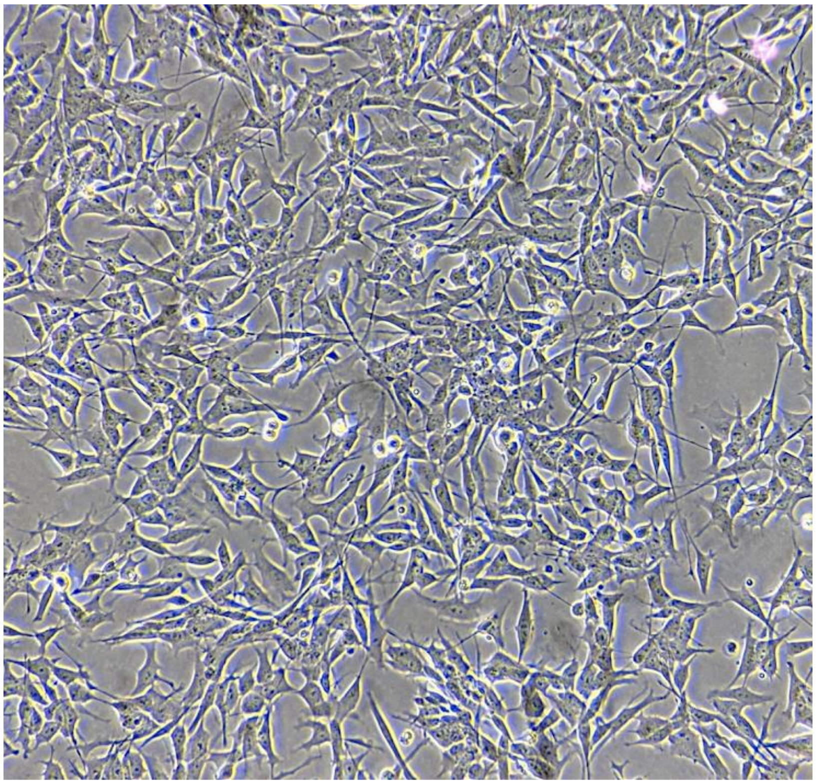


Figure 4.3 Photo of the SH-SY5Y cells prior to the pulldown assay. These cells represent ~75% confluency, the ideal confluency for an immunoprecipitation assay in general using mammalian cell lines [70-72].

4.5.3 DIG immunoprecipitation pulldown assay

In preparation for the pulldown assay, A/G protein magnetic beads (ThermoFisher Scientific™) were used as the physical pulldown method, but with anti-DIG antibodies (Jackson Immunoresearch laboratories inc.™) crosslinked to them. The cross-linking was performed by mixing the A/G magnetic beads and anti-DIG antibodies with 0.2M triethanolamine pH 8.2, and 25 mM dimethyl pimelimidate (DMP) and incubated for 45 min at room temperature. Washes were performed with 0.1M ethanolamine with 30 min incubations at room temperature. PBS washes were then performed, followed by 0.1M glycine pH 2.5 added. PBS with 0.1% TWEEN and 0.02% sodium azide washing was performed , then stored at 4°C in PBS with 0.1% TWEEN and 0.02% sodium azide.

The pulldown assays were performed in quadruplicate by lysing 1mg of human cells (for each pulldown) in 1 mL cell lysis buffer (25 mM HEPES pH 7.9, 5 mM KCl, 0.5 mM $MgCl_2$, 0.5% NP-40, 2.5 µl RiboLock™ RNase Inhibitor, 1x HALT Phosphatase and Protease Inhibitor Cocktail™ from ThermoFisher Scientific™) and adding 100µg of DIG labeled RNA that was incubated with the magnetic bead complex. In tandem, the DIG-RNA and magnetic bead complex were mixed in cell lysis buffer and incubated for 1 hour prior to mixing with the cell lysate. The RNA-magnetic bead complex was then incubated in cell lysate for 2 hours while mixing end-over-end. 3x washes were performed using 1 mL of cytoplasmic immunoprecipitation (IP) buffer (25 mM HEPES pH 7.9, 100 mM NaCl, 5 mM KCl, 0.5 mM $MgCl_2$, 0.25% NP-40, 2.5 µl RiboLock™ RNase Inhibitor, 1x HALT Phosphatase and Protease Inhibitor Cocktail). The final eluted/captured proteins were then denatured using 0.5 mL of SDS-PAGE loading dye, and run on a 12% SDS-PAGE until the samples ran 1 cm into the resolving gel in preparation for mass spectrometry analysis.

4.5.4 Mass spec and protein analysis

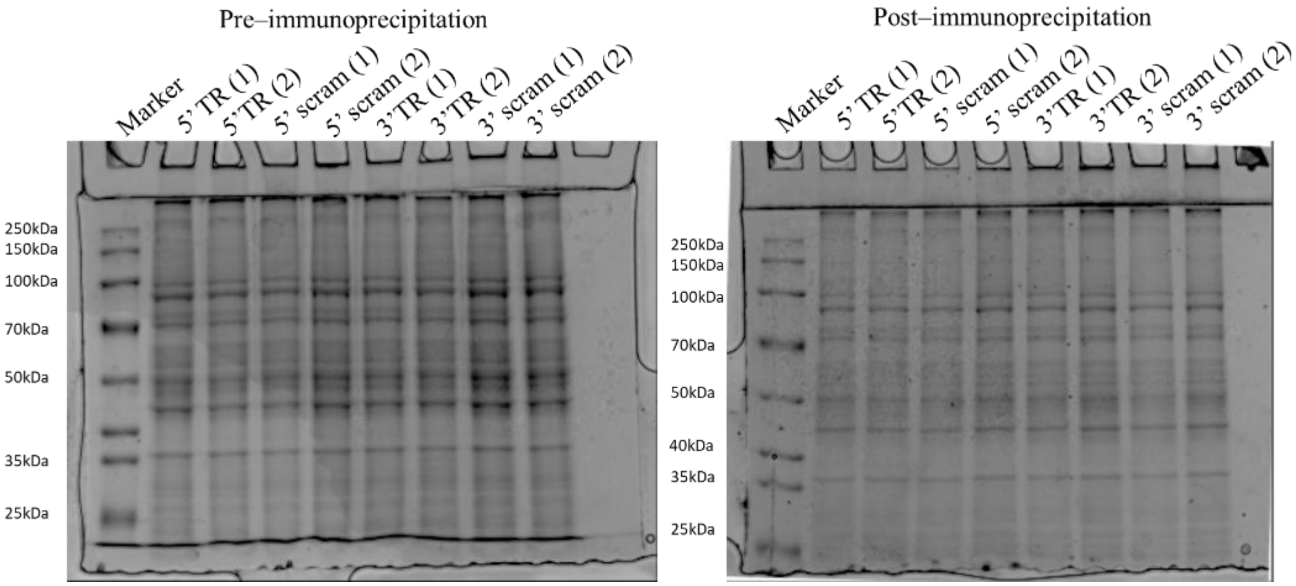
The mass spectrometry data was collected and processed by Dr. Richard Fahlman at the University of Alberta Proteomics and Mass Spectrometry Facility. In line with previous data published by the Fahlman group, each fraction sent was subjected to in-gel tryptic digestion, and the resulting peptides were dried and resuspended in 0.2% formic acid in 5% acetonitrile (ACN) [73, 74]. Digested peptides were analyzed by LC-MS/MS using a ThermoScientific™ Easy nLC-1000 in tandem with a Q-Exactive Orbitrap mass spectrometer. Each sample was resolved using a 2 cm Acclaim 100 PepMap Nanoviper C18 trapping column in tandem with a Thermo EASY-Spray column [73, 74]. Following data acquisition, raw data files were compiled for each sample and searched with Proteome Discoverer 1.4's SEQUEST search algorithm using the *Homo sapiens* complete proteome retrieved from UniProtKB [73, 74]. The non-targeted, compiled protein list was sent with their corresponding MS scores. These scores were used as part of the protein selection process, and the proteins were collectively compiled and submitted into Panther and STRING databases for elucidation of their localization and functions in the cell. Furthermore, proteins with a statistically significant E-score (a metric that uses random matches expected by chance in a database search of peptides, meaning the peptide or protein identification is more/less statistically significant and confident) as well as repetitive identification in each pulldown replicate were searched in the literature to understand if they are involved in a flaviviral infection. These collective data were used as the criterion for determined protein to invest research resources into to elucidate the functions they serve during ZIKV infection.

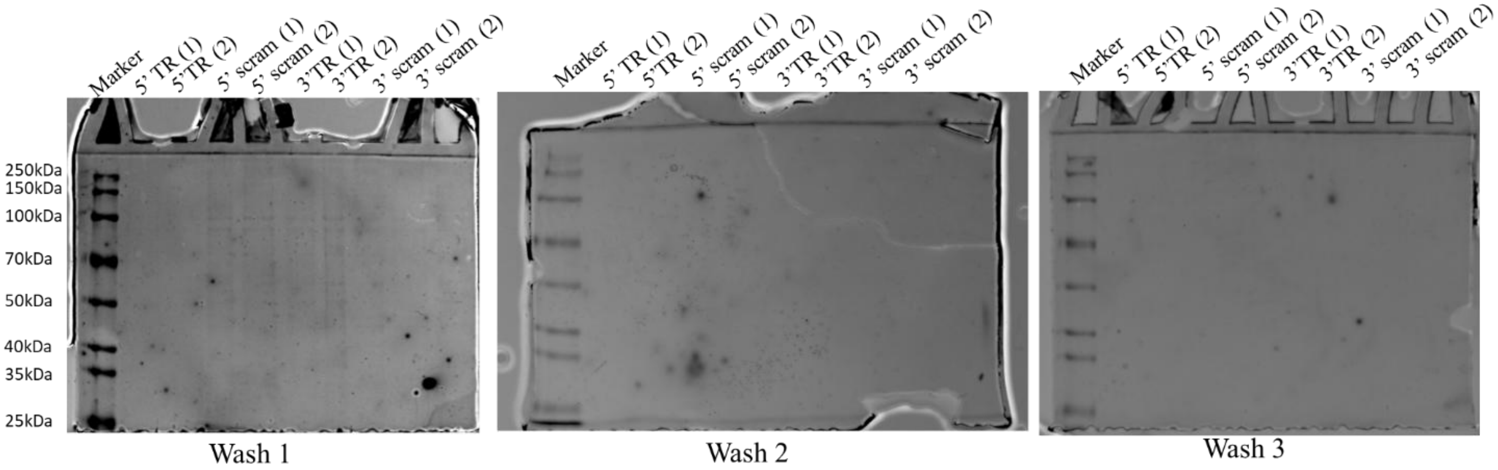
The proteins that were identified repeatedly in the pulldowns with a high E-score possessed RNA-binding capabilities, localized in the cytoplasm, and had a history of interacting with other flaviviruses were: The Interleukin-enhancer binding factors 2 & 3 (ILF2/ILF3), Nucleolin (NCL), Zinc Finger CCCH-type antiviral protein 1 (ZAP), Fragile X mental retardation syndrome-related protein 1 (FXR1), and Probable ATP-dependent RNA helicase DDX17 (DDX17). These proteins all have previous literature indicating the roles they play in flavivirus propagation. It's worth mentioning that various proteins that interact with the virus proteins or genome (dependent as well on the tissue tropism) will either confer an advantage or disadvantage for viral propagation [75-78]. Nonetheless, we investigated if these chosen proteins confer an advantage or not for ZIKV.

4.6 Results and Discussion

4.6.1 Immunoprecipitation pulldown

The pulldown assay showed significant band intensity in the pre-immunoprecipitation (cell lysate), followed by a decrease of intensity after the 2-hour incubation (Fig. 4.4). Following the incubation, 3x washes of unbound/loosely bound species were removed (Fig. 4.4), altogether, showing a decrease and disappearance in visible band intensity, suggesting that the loose/unbound species were decreasing with each subsequent wash.





Elution of binding partners

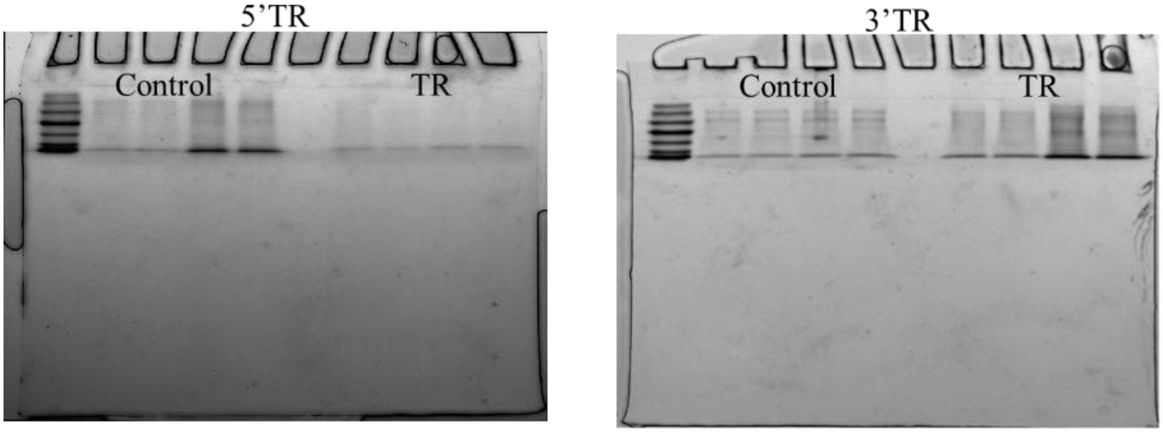


Figure 4.4 12% SDS-PAGE gels of the pre-and post-immunoprecipitation incubation, the 3 washes performed, and the elution of the binding partners of the TRs and Scrambled control RNA (Scram). Equal volumes of protein were loaded to show consistent decreasing intensity of bands from the pre-immunoprecipitation to the third wash. The elution of the binding partners were run 1 cm into the gel to keep the proteins concentrated in a smaller area to streamline mass spec identification. These elution gels were the gels sent to the Proteomics and Mass Spectrometry Facility at the University of Alberta.

Finally, the elution of the direct and indirect interacting partners (Fig. 4.4) includes not only the intensities from A/G immunoproteins and IgG antibodies used for capturing the binding partners. The most significant proteins identified in the pulldowns are contained in Table 4.1, and the proteins that bound to Scrambled RNA controls are contained in Table 4.2.

4.6.2 Binding Partner Analysis

The protein libraries identified by the Proteomics and Mass Spectrometry Facility at the University of Alberta were inputted into the PANTHER (**P**rotein **An**alysis **Th**rough **E**volutionary **R**elationships) software, which provides an inference of the various functions of genes based on published data [79]. Expanding on these functions, PANTHER allows the researcher to identify pathways, biological processes, molecular functions, and cellular locations of identified proteins [79]. Figure 4.5 demonstrates the biological processes and molecular functions of all identified proteins in the pulldowns. The results indicate that the proteins identified in the 5′ and 3′ TR pulldowns have similar biological functions such as developmental, immune system, and localization processes, as well as a general cellular function (Fig. 4.5). Furthermore, when analyzing the molecular functions, the predominant function of these proteins are nucleic acid binding proteins, which is one indicator that the pulldown assay was successful at capturing nucleic acid binding proteins (Fig. 4.5). Previous evidence shows the ZIKV 3′ TR is involved in suppression of the immune response by suppressing host mRNA turnover including cytokine mRNA [80]. Furthermore, PANTHER showed developmental proteins binding to the TRs, which has been previously demonstrated with the Fragile X mental retardation protein [81]. This protein is critical for the neurological development of a fetus, and it was previously shown that the 3′ TR hijacks this protein, creating conditions for neurological sequelae [81]. Notably, this protein was identified in the pulldown assays for the 3′ TR. Altogether, all ~148 proteins identified were collected and inputted into the PANTHER cellular and disease-related pathways and showed 16 pathways that the TR binding partners are involved in based on previous literature (Fig. 4.7). These include Alzheimer’s, Parkinson’s, Huntington’s neurological pathologies, but also, cytokine, inflammation, and apoptosis signalling proteins. Moreover, these identified binding partners involved in these pathways were considered in choosing the proteins to go forward with in characterization with the TRs.

Table 4.1 Significant binding partners identified from the mass spec results of the ZIKV 5' and 3' TR pulldown assays. The best expected value (Best Expect Val) collects all the mass spec data and gives the researcher a confidence value that the protein identified by the software is that specific protein. Identified proteins labeled blue indicate the selected proteins that were further validated and studied. Reiterating, these were chosen based on their score, ease of purification, and previous published data that they were present in flaviviral infection. The red DHX9 was not selected due to previously published data indicating that DHX9 had no influence on ZIKV replication when knocked down [82]. All blank cells indicate an absence of that protein in the specific pulldown performed (eg. FXR1 was not identified in the 3' TR pulldowns 3 & 4).

ZIKV 5' TR Pulldown 1	**ZIKV 5' TR Pulldown 2**	**ZIKV 5' TR Pulldown 3**	**ZIKV 5' TR pulldown 4**	
Best Expect Val	Best Expect Val	Best Expect Val	Best Expect Val	**Selection**
8.90E-08	9.70E-07	6.80E-07	3.40E-05	DHX9
5.10E-04	9.80E-05	7.90E-06	5.00E-05	ILF3
1.20E-06	5.70E-05			ILF2
4.20E-04	1.60E-05			DHX30
	4.50E-04	0.0025	3.20E-04	Nucleolin
		3.00E-06	1.30E-09	Zinc finger CCCH type antiviral protein1
2.50E-06	7.00E-05			DDX5

ZIKV 3' TR Pulldown 1	**ZIKV 3' TR pulldown 2**	**ZIKV 3' TR pulldown 3**	**ZIKV 3' TR Pulldown 4**	
Best Expect Val	Best Expect Val	Best Expect Val	Best Expect Val	**Selection**
6.70E-05	8.30E-06	5.80E-08	1.40E-09	DHX9
4.10E-07	4.20E-07		4.50E-08	ILF2
1.30E-05	2.90E-05			FXR1
5.20E-04	1.60E-05		3.00E-07	Splicing factor 3
3.90E-04	3.40E-05	3.20E-09	2.60E-08	Nucleolin
			1.30E-06	DDX17
8.50E-06	0.0011	0.0031	1.20E-05	ILF3
8.80E-06	2.10E-06			Splicing factor, proline- and glutamine-rich
1.40E-05	0.0017	1.70E-07	1.80E-07	Nuclease-sensitive element-binding protein 1

0.0013	7.20E-04		5.70E-05	RNA-binding protein FUS
1.80E-04	1.60E-04		8.10E-07	VGF neurosecretory protein
4.40E-07	9.80E-07	4.70E-04	2.30E-06	Insulin-like growth factor 2 mRNA-binding protein 3
5.20E-05	4.60E-06		1.40E-04	Leucine-rich repeat-containing protein 59
			1.40E-07	Zinc finger CCCH type antiviral protein 1
8.50E-06	1.40E-05			DDX5

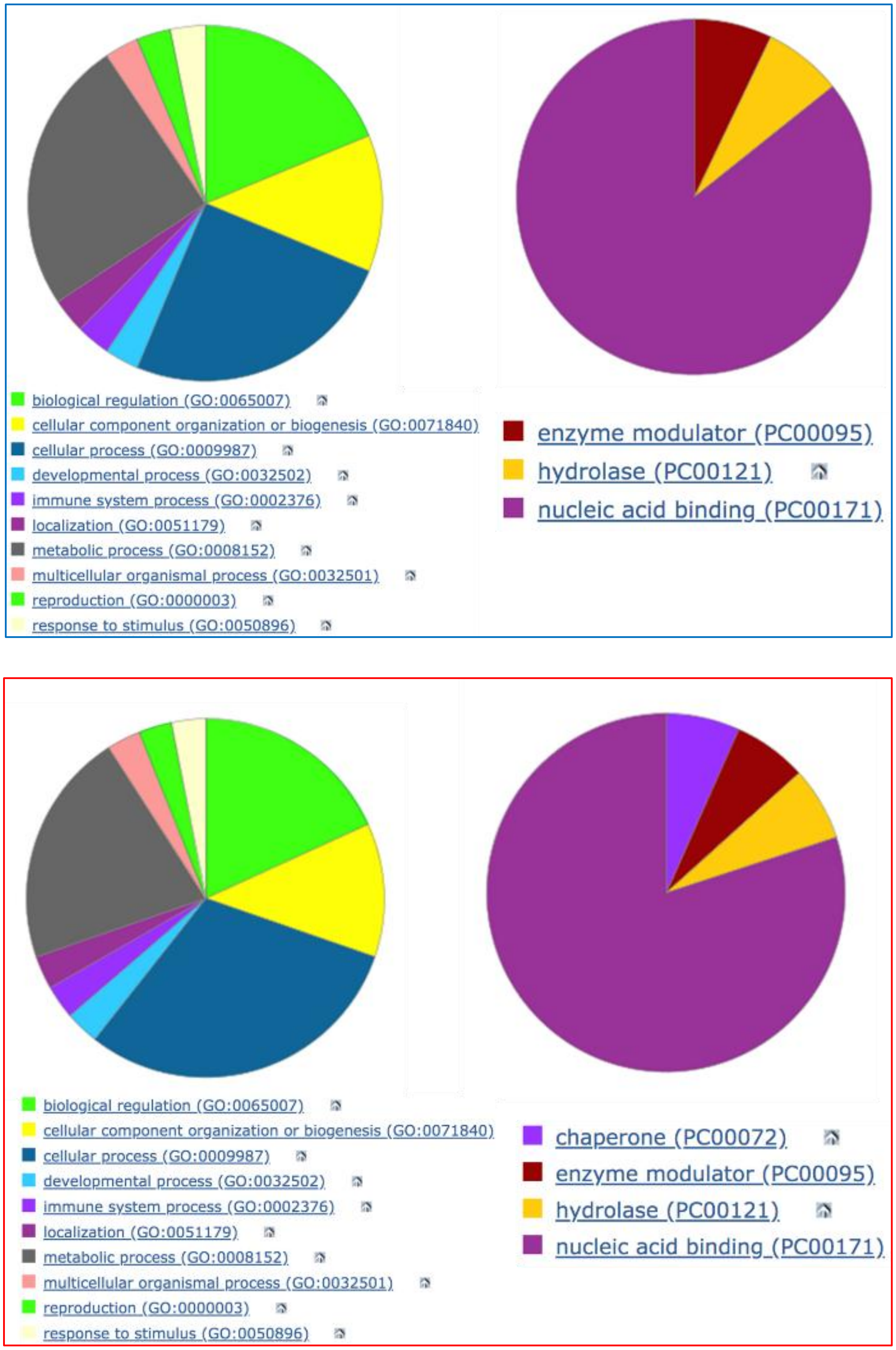
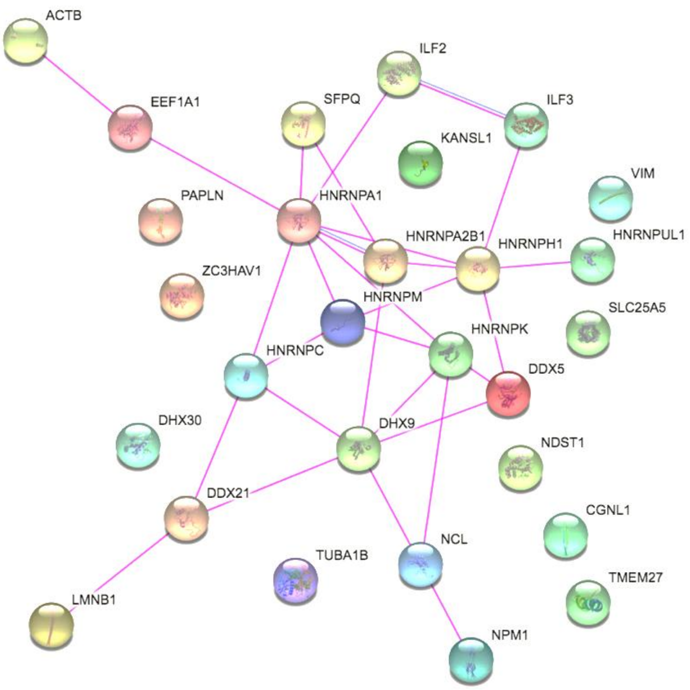


Figure 4.5 PANTHER analysis of the ZIKV 5' (blue outlined chart) and 3' (red outlined chart) TR pulldown assay binding partners and their respective biological and specific molecular functions. The left chart in each indicates the overall distribution of biological processes that these proteins participate in that PANTHER could identify, and the right shows the molecular functions of identified proteins.

Following the PANTHER analysis, the 5’ and 3’ TR binding partners were inputted into the STRING (**S**earch **T**ool for the **R**etrieval of **In**teracting **G**enes/Proteins) suite as well. This program allows for the detection of protein-protein interaction networks of inputted proteins (Fig. 4.6) [83]. The software detects protein-protein interactions by using; Experimentally determined, curated databases, gene neighborhood, gene fusions, gene co-occurrence, text mining, co-expression, and protein homology [83]. These options have their respective place depending on the researchers’ goals, however, for our purposes, we selected only experimentally determined and curated databases. Curated databases regarding STRING mean that the protein-protein interaction data obtained is from scientific literature and experimental studies, giving further credence that the interactions shown are credible. Figure 4.6 below shows the protein-protein networks of the individual 5’ and 3’ TR binding partners. These results allow us to determine not only if the proteins identified are directly/indirectly interacting with the TRs, but also, provide insight into the intricate pathways that these proteins could be involved in, thus paving the way for potential therapeutic targets.


ACTB
ILF2
EEF1A1
SFPQ
ILF3
KANSL1
VIM
PAPLN
HNRNPA1
HNRNPA2B1
HNRNPH1
HNRNPUL1
ZC3HAV1
HNRNPM
SLC25A5
HNRNPK
HNRNPC
DDX5
DHX30
DHX9
NDST1
DDX21
CGNL1
NCL
TUBA1B
TMEM27
LMNB1
NPM1

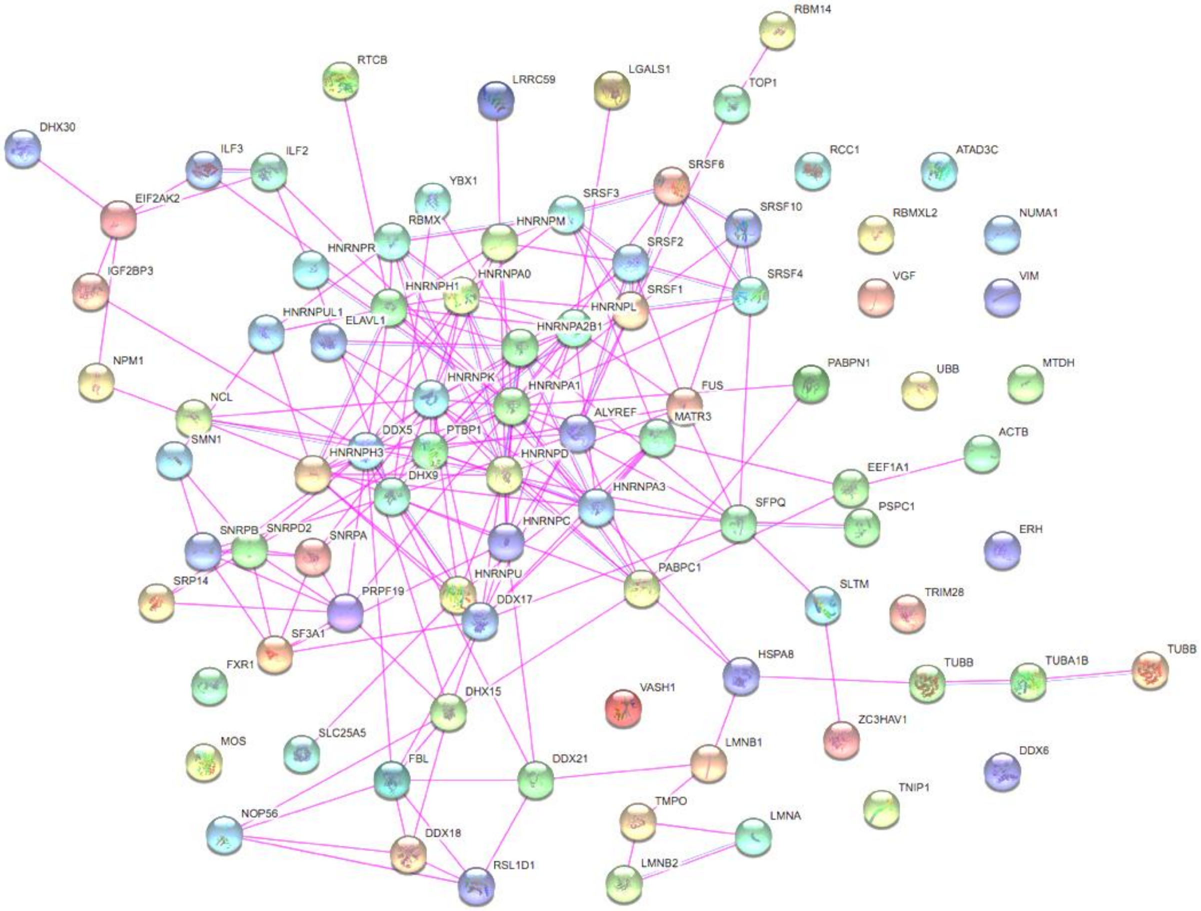

RBM14
RTCB
LRRC59
LGALS1
TOP1
DHX30
ILF3
ILF2
RCC1
ATAD3C
SRSF6
YBX1
SRSF3
EIF2AK2
RBMX
HNRNPM
SRSF10
NUMA1
RBMXL2
HNRNPR
SRSF2
IGF2BP3
HNRNPH1
HNRNPA0
SRSF4
VGF
VIM
SRSF1
HNRNPL
HNRNPUL1
ELAVL1
HNRNPA2B1
NPM1
PABPN1
UBB
MTDH
HNRNPK
NCL
HNRNPA1
FUS
ALYREF
MATR3
SMN1
DDX5
PTBP1
ACTB
HNRNPH3
HNRNPD
DHX9
EEF1A1
HNRNPA3
PSPC1
SFPQ
SNRPB
SNRPD2
SNRPA
HNRNPC
ERH
PABPC1
SRP14
HNRNPU
PRPF19
DDX17
SLTM
TRIM28
SF3A1
FXR1
HSPA8
TUBB
TUBA1B
TUBB
VASH1
DHX15
ZC3HAV1
MOS
SLC25A5
LMNB1
FBL
DDX21
DDX6
TMPO
TNIP1
NOP56
LMNA
DDX18
RSL1D1
LMNB2

Figure 4.6. Raw STRING analysis of all protein-protein interactions and networks identified in the pulldown assay that STRING could identify. The above schematic is the 5’ TR (28 proteins) and the bottom is the 3’ TR (~120 proteins). Using a medium confidence filter, as well as showing only protein-protein interactions identified in experiments (lavender line) and Curated databases (pink line), the networks represent associations that these proteins make with each other, altogether aiding in the protein selection, as well as hypothesizing if the proteins identified are directly or indirectly interacting with the viral RNA.

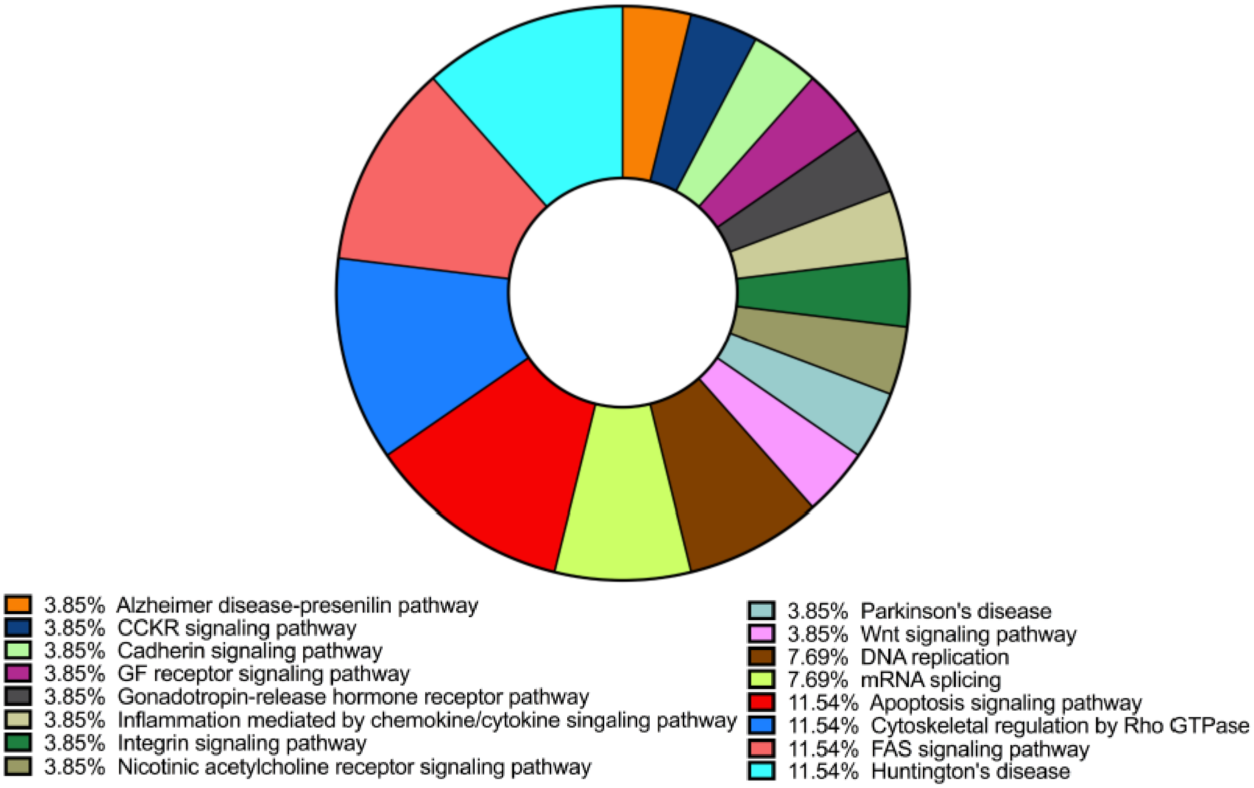


Figure 4.7 ~148 proteins were identified in the pulldown altogether for the 5’ and 3’ TR pulldowns, and using the PANTHER suite, sixteen cellular and disease-related pathways were identified wherein the proteins were involved. Above is a pie chart which highlights the pathways where each of the ~148 proteins are involved. 6 top-ranking proteins were chosen for expression, purification, and characterization with the ZIKV TRs. The percentages are a representation of the frequency at which the identified proteins show up in the respective pathways. Furthermore, the percentages are only proteins that were identified in these pathways overall and not a percentage of the overall binding partners identified in the pulldowns.

Table 4.2 All proteins identified in the pulldowns that were found to bind to scrambled RNAs. This control was to give insight into general direct and indirect RNA binding proteins and assist in deciding which proteins to focus on for downstream studies. Note that binding of general RNA binding proteins does not negate criticality of its role in ZIKV replication, hence why some of the proteins in the list below were still chosen for downstream characterization. 3 groups of tables are below - Unique to 5′ Scrambled RNAs, Unique to 3′ Scrambled RNAs, and Duplicates (present in both scrambled RNAs).

Unique to 5′ Scrambled RNAs
Heterogeneous nuclear ribonucleoprotein A/B
Actin, cytoplasmic 1
Dentin sialophosphoprotein
Heterogeneous nuclear ribonucleoprotein K
Kinesin heavy chain isoform 5C
V(D)J recombination-activating protein 2
Ketohexokinase
Metallothionein-1B
Forkhead box protein D3
Fibronectin

Unique to 3′ Scrambled RNAs
Heterogeneous nuclear ribonucleoprotein Q
60S ribosomal protein L4

Duplicates (present in both scrambled RNAs)
40S ribosomal protein S7
Heterogeneous nuclear ribonucleoprotein H
40S ribosomal protein S16
Heterogeneous nuclear ribonucleoprotein D-like
40S ribosomal protein S10
RNA-binding protein 28
40S ribosomal protein SA
Serine/arginine-rich splicing factor 3
Paraspeckle component 1
Serine/arginine-rich splicing factor 7
Protein FAM3B
Zinc finger protein 536
LON peptidase N-terminal domain and RING finger protein 3
Putative inactive cytochrome P450 family member 4Z2
Zonadhesin
Leucine-rich repeat-containing protein 14

4.6.3 ZIKV TR binding partner selection for downstream experiments

The selection criteria for moving forward with downstream characterization with the TRs was based on: the mass spec scores and the frequency of 'hits' in the quadruplicate replicates in the pulldowns, their ease of expression and purification based on previous literature, if the proteins have a history of interacting with flaviviruses or viruses in general based on database searches, and if the proteins were involved in any neurological or developmental diseases based on the PANTHER results.

4.6.4 ILF2 & ILF3

ILF2/3 were discovered in the pulldown to bind to both the 5' and 3' TRs. ILF2 & 3 form a strongly bound heterodimer and are almost exclusively seen interacting with each other [84-87]. ILF2 & 3 can act as a positive or negative regulators of gene expression, act as an IRES trans-acting factor that is activated during stress response to protein unfolding, and are involved in complexing together and regulating mitotic gene expression by competing with Staufen-mediated mRNA decay [84-87]. Regarding ILF2/3's involvement in flaviviral replication, it's been shown that they interact with hepatitis C and aid in replication [88, 89]. It was also shown that ILF3 interacts with the 3' SL of the 3' TR of DENV and enhances viral replication. Knockdown of ILF3 decreased the RNA transcript levels by 50-70% [90]. Conversely, it was shown *in vitro* that ILF2 possesses antiviral activity against the Japanese Encephalitis virus (JEV) [91]. It's worth noting that they did not determine where it was bound on JEV [91]. The overexpression in *E. coli* we carried forward with ILF2/3 was attempted with a co-expression and affinity purification. Previously, a research group has crystalized the ILF2/3 complex, albeit with a 50% truncation of ILF3, using the *Mus* *musculus* ILF2/3 proteins [92]. The double-stranded RNA binding domains 1 & 2 of this crystal structure were removed; therefore, the structure is not relevant to our research group investigating potential RNA binding capabilities. It was also discovered that a family of non-coding RNAs that so far have been discovered to exclusively bind to ILF3 in humans called, snaR (small NF90 (ILF3) associated RNA), might be worth investigating, as the function of these RNAs remains unknown, and if these RNA are interacting with ILF3, it stands to reason that it could be binding to ZIKV during replication [93-95]. We aim to investigate if ILF2/3 are binding directly to ZIKV RNA, participate in aiding viral replication, and if snaR RNA is potentially binding to ZIKV RNA, and ILF3 is either acting as a protein with snaR being a 'guide RNA' and providing any further functionality to this interaction.

4.6.5 Nucleolin

NCL was shown to be present in all pulldowns for both the 5' and 3' TRs. NCL is a multifunctional protein that is involved in ribosome biogenesis and maturation, gene silencing, senescence, cell cycle regulation, modulating proliferation and apoptosis, and nucleo-cytoplasmic transport of RNA [96-100]. Previous literature has shown that NCL interacts with DENV capsid protein and viral RNA, and plays a role in viral packaging [101]. The research group showed that the knockdown of NCL in HEK293 cells resulted in a significant decrease in viral titers. Furthermore, another study showed that NCL bound and stabilized a G-quadruplex in the core gene of HCV, and the knockdown of NCL resulted in an enhancement in viral RNA replication [102]. The group hypothesizes that NCL may function as a tool in the innate immunity for HCV by stabilizing the G-quadruplex [102]. Due to the previous literature on NCL and flaviviruses, it is unclear if NCL confers an advantage or not for ZIKV. We characterized the interactions NCL makes with ZIKV TRs and show that it is critical for viral replication by performing knockdowns using a ZIKV replicon system, and as well, in collaboration with the Chatel-Chaix lab in Quebec, show that viral titers and viral RNA transcripts are significantly decreased (see Chapter 5).

4.6.6 Zinc Finger CCCH-type antiviral protein 1

The pulldown assays revealed binding occurring from the ZAP protein interacting with both TRs. ZAP has been shown in previous literature to be involved in restricting viral replication, relieving immunosuppression, positively regulating type I interferon response, interacting with viral RNA, chromatin remodelling, and DNA repair [103-106]. There were two studies performed on ZAP interacting with the flaviviruses, and both were performed using JEV. One of them reports that ZAP hindered viral translation, and as well, targeted the viral RNA for 3’-5’ exosome-mediated degradation, and showed that knockdown of ZAP in human neuronal BE(2)C cells enhanced JEV replication [104]. The other study showed the same translational inhibition, but also, showed that targeting the degradation is a consequence of CpG dinucleotides, which flaviviruses have a bias towards decreasing, and it’s hypothesized that it’s to avoid being signalled for degradation [107-109]. Due to the antiviral activity that ZAP possesses on JEV, it’s unknown if it is a disadvantage for ZIKV replication, as it could be an advantage. Further investigation will be performed.

4.6.7 Fragile X mental retardation syndrome-related protein 1

FXR1 was a protein that was identified to interact with the 3’ TR in the pulldowns. This protein is highly expressed in the brain and reproductive tissues and is directly involved in the autism spectrum disorder, as well as amyotrophic lateral sclerosis [110, 111]. The protein is involved in translational control by binding to RNA, specifically, RNA G-quadruplexes/G-quartets, and neuronal function by aiding in synaptic plasticity, axonal guidance, and dendritic spine morphology [111-114]. When this protein interacts with RNA it induces liquid-liquid phase separation, which is altogether regulated by FXR1 phosphorylation [115, 116]. In general, phase separation is a process wherein typically non-coding RNA recruits proteins as a molecular scaffold, which forms multivalent interactions within a droplet condensate in the cell [117, 118]. This can isolate certain ions from the vesicle and control the binding efficiency by keeping certain biomolecules in the vesicle and proximity allowing for more binding events to occur [119, 120]. Furthermore, there is literature on FMRP, a homolog of FXR1 interacting with the ZIKV 3’ TR [81]. The group showed that FXR1 interacts with the sfRNA of ZIKV and could be involved in suppressing FXR1. Knockdown of FXR1 in HeLa cells increased viral replication, showing that it is a restriction factor for viral replication, and the sfRNA antagonizes FXR1 activity [81]. We aimed to biophysically characterize this interaction with the 3’ TR and investigate if we have a similar result if we knock down FXR1 in SH-SY5Y cells.

4.6.8 Probable ATP-dependent RNA helicase DDX17

DDX17 was identified in the pulldowns to specifically interact with the 3’ TR. This protein is involved in pre-mRNA splicing, alternative splicing, rRNA processing, miRNA processing transcriptional regulation, and modulation of RNA function [121-125]. Regarding flaviviruses and their interactions with DDX17, it was discovered to interact with tick-borne encephalitis RNA [126]. Moreover, we identify that the ZIKV 3’ TR contains a conserved non-canonical structure known as a G-quadruplex, and we show that DDX17 unfolds this region (see Chapter 6). Furthermore, it is unclear what the specific function(s) of DDX17 are during ZIKV replication, and we aim to identify those functions and elucidate if DDX17 is critical for viral replication or not.

4.7 Conclusion

Immunoprecipitation pulldown assays are a vital experiment for determining binding partners that interact with a specific ‘bait’ that the researcher is studying and remains a strong tool in identification techniques. For our research goals, using DIG-labelled ZIKV 5’ and 3’ TRs as bait we were able to identify potential protein binding partners using cell lysate from the SH-SY5Y neuronal-like cell line. We chose five proteins to carry forward in validating the pulldown assay, characterizing the interaction itself, and validating if the proteins are critical for ZIKV replication. A considerable amount of research is still required to investigate if these proteins can be exploited for therapeutic purposes to combat ZIKV, but using a basic science approach, we provide a foundational understanding of the interactions the TRs make with these proteins.

4.8 References

1. Heiny, A.T., et al., *Evolutionarily Conserved Protein Sequences of Influenza A Viruses, Avian and Human, as Vaccine Targets.* PLOS ONE, 2007. **2**(11): p. e1190.
2. ElHefnawi, M., et al., *Identification of novel conserved functional motifs across most Influenza A viral strains.* Virology Journal, 2011. **8**(1): p. 44.
3. Marsh, G.A., et al., *Highly conserved regions of influenza a virus polymerase gene segments are critical for efficient viral RNA packaging.* J Virol, 2008. **82**(5): p. 2295-304.
4. Rodriguez, P., et al., *Mutations of the segment-specific nucleotides at the 3' end of influenza virus NS segment control viral replication.* Virology, 2020. **539**: p. 104-113.
5. Petersen, H., et al., *NS Segment of a 1918 Influenza A Virus-Descendent Enhances Replication of H1N1pdm09 and Virus-Induced Cellular Immune Response in Mammalian and Avian Systems.* Frontiers in Microbiology, 2018. **9**.
6. Bogdanow, B., et al., *The dynamic proteome of influenza A virus infection identifies M segment splicing as a host range determinant.* Nature Communications, 2019. **10**(1): p. 5518.
7. Marsh, G.A., et al., *Highly Conserved Regions of Influenza A Virus Polymerase Gene Segments Are Critical for Efficient Viral RNA Packaging.* Journal of Virology, 2008. **82**(5): p. 2295-2304.
8. Duan, Y., et al., *Flavivirus RNA-Dependent RNA Polymerase Interacts with Genome UTRs and Viral Proteins to Facilitate Flavivirus RNA Replication.* Viruses, 2019. **11**(10).
9. Ferrer-Orta, C., D. Ferrero, and N. Verdaguer, *RNA-Dependent RNA Polymerases of Picornaviruses: From the Structure to Regulatory Mechanisms.* Viruses, 2015. **7**(8): p. 4438-60.
10. Martin, R., et al., *Genetic conservation of SARS-CoV-2 RNA replication complex in globally circulating isolates and recently emerged variants from humans and minks suggests minimal pre-existing resistance to remdesivir.* Antiviral Res, 2021. **188**: p. 105033.
11. Zhou, Z., et al., *Influenza A virus polymerase: an attractive target for next-generation anti-influenza therapeutics.* Drug Discovery Today, 2018. **23**(3): p. 503-518.
12. Duprex, W.P., F.M. Collins, and B.K. Rima, *Modulating the function of the measles virus RNA-dependent RNA polymerase by insertion of green fluorescent protein into the open reading frame.* J Virol, 2002. **76**(14): p. 7322-8.
13. Kiriwan, D. and K. Choowongkomon, *In silico structural elucidation of the rabies RNA-dependent RNA polymerase (RdRp) toward the identification of potential rabies virus inhibitors.* J Mol Model, 2021. **27**(6): p. 183.
14. McDonald, S.M., Y.J. Tao, and J.T. Patton, *The ins and outs of four-tunneled Reoviridae RNA-dependent RNA polymerases.* Curr Opin Struct Biol, 2009. **19**(6): p. 775-82.
15. Jenni, S., et al., *In situ Structure of Rotavirus VP1 RNA-Dependent RNA Polymerase.* J Mol Biol, 2019. **431**(17): p. 3124-3138.
16. Baker, A.C. and D.C. Schroeder, *The use of RNA-dependent RNA polymerase for the taxonomic assignment of Picorna-like viruses (order Picornavirales) infecting Apis mellifera L. populations.* Virology Journal, 2008. **5**(1): p. 10.
17. Duflos, C. and T. Michiels, *Regulation of viral RNA-dependent RNA polymerases by phosphorylation.* Frontiers in Virology, 2023. **3**.

18. Gao, Y., et al., *Structure of the RNA-dependent RNA polymerase from COVID-19 virus.* Science, 2020. **368**(6492): p. 779-782.
19. Bruenn, J.A., *A structural and primary sequence comparison of the viral RNA-dependent RNA polymerases.* Nucleic Acids Res, 2003. **31**(7): p. 1821-9.
20. Choi, K.H., *Viral polymerases.* Adv Exp Med Biol, 2012. **726**: p. 267-304.
21. Logan, G., et al., *Deep Sequencing of Foot-and-Mouth Disease Virus Reveals RNA Sequences Involved in Genome Packaging.* J Virol, 2018. **92**(1).
22. Hagey, R.J., et al., *Programmable antivirals targeting critical conserved viral RNA secondary structures from influenza A virus and SARS-CoV-2.* Nature Medicine, 2022. **28**(9): p. 1944-1955.
23. Thomas, S.J. and I.K. Yoon, *A review of Dengvaxia®: development to deployment.* Hum Vaccin Immunother, 2019. **15**(10): p. 2295-2314.
24. Barton, M.A. and M.I. Salvadori, *Zika virus and microcephaly.* Cmaj, 2016. **188**(7): p. E118-e119.
25. Wen, Z., H. Song, and G.L. Ming, *How does Zika virus cause microcephaly?* Genes Dev, 2017. **31**(9): p. 849-861.
26. Organization, W.H. *Zika Virus*. 2022.
27. Kolimenakis, A., et al., *The role of urbanisation in the spread of Aedes mosquitoes and the diseases they transmit-A systematic review.* PLoS Negl Trop Dis, 2021. **15**(9): p. e0009631.
28. Kulkarni, M.A., C. Duguay, and K. Ost, *Charting the evidence for climate change impacts on the global spread of malaria and dengue and adaptive responses: a scoping review of reviews.* Globalization and Health, 2022. **18**(1): p. 1.
29. Vora, N., *Impact of anthropogenic environmental alterations on vector-borne diseases.* Medscape J Med, 2008. **10**(10): p. 238.
30. Crowder, D.W., et al., *West nile virus prevalence across landscapes is mediated by local effects of agriculture on vector and host communities.* PLoS One, 2013. **8**(1): p. e55006.
31. Kausar, S., et al., *A review: Mechanism of action of antiviral drugs.* Int J Immunopathol Pharmacol, 2021. **35**: p. 20587384211002621.
32. Koike, H. and M. Katsuno, *Emerging infectious diseases, vaccines and Guillain-Barré syndrome.* Clin Exp Neuroimmunol, 2021. **12**(3): p. 165-170.
33. Baz, M. and G. Boivin, *Antiviral Agents in Development for Zika Virus Infections.* Pharmaceuticals (Basel), 2019. **12**(3).
34. Fong, Y.D. and J.J.H. Chu, *Natural products as Zika antivirals.* Med Res Rev, 2022. **42**(5): p. 1739-1780.
35. Carvalho, D.C.M., et al., *Antiviral activity of ouabain against a Brazilian Zika virus strain.* Scientific Reports, 2022. **12**(1): p. 12598.
36. Saiz, J.-C., *Therapeutic Advances Against ZIKV: A Quick Response, a Long Way to Go.* Pharmaceuticals, 2019. **12**(3): p. 127.
37. Mahajan, S., et al., *Antiviral strategies targeting host factors and mechanisms obliging +ssRNA viral pathogens.* Bioorg Med Chem, 2021. **46**: p. 116356.
38. Chen, N., et al., *Virus-host interaction networks as new antiviral drug targets for IAV and SARS-CoV-2.* Emerg Microbes Infect, 2022. **11**(1): p. 1371-1389.
39. Roa-Linares, V.C., et al., *Host Cell Targets for Unconventional Antivirals against RNA Viruses.* Viruses, 2023. **15**(3): p. 776.

40. de Chassey, B., et al., *Virus-host interactomics: new insights and opportunities for antiviral drug discovery.* Genome Medicine, 2014. **6**(11): p. 115.
41. Cakir, M., et al., *Target Discovery for Host-Directed Antiviral Therapies: Application of Proteomics Approaches.* mSystems, 2021. **6**(5): p. 10.1128/msystems.00388-21.
42. Badia, R., E. Garcia-Vidal, and E. Ballana, *Viral-Host Dependency Factors as Therapeutic Targets to Overcome Antiviral Drug-Resistance: A Focus on Innate Immune Modulation.* Frontiers in Virology, 2022. **2**.
43. Furió, V., A. Moya, and R. Sanjuán, *The cost of replication fidelity in an RNA virus.* Proceedings of the National Academy of Sciences, 2005. **102**(29): p. 10233-10237.
44. Kautz, T.F. and N.L. Forrester, *RNA Virus Fidelity Mutants: A Useful Tool for Evolutionary Biology or a Complex Challenge?* Viruses, 2018. **10**(11).
45. Castro, C., J.J. Arnold, and C.E. Cameron, *Incorporation fidelity of the viral RNA-dependent RNA polymerase: a kinetic, thermodynamic and structural perspective.* Virus Res, 2005. **107**(2): p. 141-9.
46. Campagnola, G., et al., *Structure-Function Relationships Underlying the Replication Fidelity of Viral RNA-Dependent RNA Polymerases.* Journal of Virology, 2015. **89**(1): p. 275-286.
47. Smith, E.C., N.R. Sexton, and M.R. Denison, *Thinking Outside the Triangle: Replication Fidelity of the Largest RNA Viruses.* Annual Review of Virology, 2014. **1**(1): p. 111-132.
48. Brinton, M.A. and M. Basu, *Functions of the 3' and 5' genome RNA regions of members of the genus Flavivirus.* Virus Res, 2015. **206**: p. 108-19.
49. Ng, W.C., et al., *The 5' and 3' Untranslated Regions of the Flaviviral Genome.* Viruses, 2017. **9**(6).
50. Berzal-Herranz, A., et al., *The Genomic 3' UTR of Flaviviruses Is a Translation Initiation Enhancer.* Int J Mol Sci, 2022. **23**(15).
51. Slonchak, A., et al., *Structural analysis of 3’UTRs in insect flaviviruses reveals novel determinants of sfRNA biogenesis and provides new insights into flavivirus evolution.* Nature Communications, 2022. **13**(1): p. 1279.
52. Diosa-Toro, M., et al., *Role of RNA-binding proteins during the late stages of Flavivirus replication cycle.* Virology Journal, 2020. **17**(1): p. 60.
53. Borba, L.d., et al., *RNA Structure Duplication in the Dengue Virus 3’ UTR: Redundancy or Host Specificity?* mBio, 2019. **10**(1): p. 10.1128/mbio.02506-18.
54. Bavia, L., et al., *A glance at subgenomic flavivirus RNAs and microRNAs in flavivirus infections.* Virology Journal, 2016. **13**(1): p. 84.
55. Bonenfant, G., et al., *Zika Virus Subverts Stress Granules To Promote and Restrict Viral Gene Expression.* Journal of Virology, 2019. **93**(12): p. 10.1128/jvi.00520-19.
56. Soto-Acosta, R., et al., *Fragile X mental retardation protein is a Zika virus restriction factor that is antagonized by subgenomic flaviviral RNA.* eLife, 2018. **7**: p. e39023.
57. Chen, X., et al., *Zika virus RNA structure controls its unique neurotropism by bipartite binding to Musashi-1.* Nature Communications, 2023. **14**(1): p. 1134.
58. Michalski, D., et al., *Zika virus noncoding sfRNAs sequester multiple host-derived RNA-binding proteins and modulate mRNA decay and splicing during infection.* Journal of Biological Chemistry, 2019. **294**(44): p. 16282-16296.

59. Stothard, P., *The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences.* Biotechniques, 2000. **28**(6): p. 1102, 1104.
60. Tetin, S.Y., K.M. Swift, and E.D. Matayoshi, *Measuring antibody affinity and performing immunoassay at the single molecule level.* Analytical Biochemistry, 2002. **307**(1): p. 84-91.
61. Gemmill, D., et al., *Current approaches for RNA-labelling to identify RNA-binding proteins.* Biochemistry and Cell Biology, 2020. **98**(1): p. 31-41.
62. Schneider, C.A., W.S. Rasband, and K.W. Eliceiri, *NIH Image to ImageJ: 25 years of image analysis.* Nature Methods, 2012. **9**(7): p. 671-675.
63. Xicoy, H., B. Wieringa, and G.J.M. Martens, *The SH-SY5Y cell line in Parkinson's disease research: a systematic review.* Molecular Neurodegeneration, 2017. **12**(1): p. 10.
64. Hoffmann, L.F., et al., *Neural regeneration research model to be explored: SH-SY5Y human neuroblastoma cells.* Neural Regen Res, 2023. **18**(6): p. 1265-1266.
65. Alrashidi, H., S. Eaton, and S. Heales, *Biochemical characterization of proliferative and differentiated SH-SY5Y cell line as a model for Parkinson's disease.* Neurochemistry International, 2021. **145**: p. 105009.
66. Cai, A., et al., *Neuroblastoma SH-SY5Y Cell Differentiation to Mature Neuron by AM580 Treatment.* Neurochemical Research, 2022. **47**(12): p. 3723-3732.
67. Gangras, P., et al., *Investigating SH-SY5Y Neuroblastoma Cell Surfaceome as a Model for Neuronal-Targeted Novel Therapeutic Modalities.* International Journal of Molecular Sciences, 2022. **23**(23): p. 15062.
68. Sánchez-San Martín, C., et al., *Differentiation enhances Zika virus infection of neuronal brain cells.* Sci Rep, 2018. **8**(1): p. 14543.
69. Pepe, A., et al., *Tunneling nanotubes provide a route for SARS-CoV-2 spreading.* Science Advances, 2022. **8**(29): p. eabo0171.
70. Hsu, J.C., et al., *Protocol for assessing translational regulation in mammalian cell lines by OP-Puro labeling.* STAR Protoc, 2022. **3**(3): p. 101654.
71. Cattoglio, C., et al., *Assessing Self-interaction of Mammalian Nuclear Proteins by Co-immunoprecipitation.* Bio Protoc, 2020. **10**(4): p. e3526.
72. Lagundžin, D., et al., *An optimized co-immunoprecipitation protocol for the analysis of endogenous protein-protein interactions in cell lines using mass spectrometry.* STAR Protocols, 2022. **3**(1): p. 101234.
73. Piragasam, R.S., et al., *Label-free proteomic analysis reveals large dynamic changes to the cellular proteome upon expression of the miRNA-23a-27a-24-2 microRNA cluster.* Biochem Cell Biol, 2020. **98**(1): p. 61-69.
74. Kramer, D.A., et al., *Proteomic characterization of EL4 lymphoma-derived tumors upon chemotherapy treatment reveals potential roles for lysosomes and caspase-6 during tumor cell death in vivo.* Proteomics, 2017. **17**(12).
75. Meier-Stephenson, V., et al., *DEAD-box helicases: the Yin and Yang roles in viral infections.* Biotechnol Genet Eng Rev, 2018. **34**(1): p. 3-32.
76. Jones, J.E., V. Le Sage, and S.S. Lakdawala, *Viral and host heterogeneity and their effects on the viral life cycle.* Nature Reviews Microbiology, 2021. **19**(4): p. 272-282.
77. Goodacre, N., et al., *Protein-protein interactions of human viruses.* Semin Cell Dev Biol, 2020. **99**: p. 31-39.

78. Villanueva, R.A., Y. Rouillé, and J. Dubuisson, *Interactions between virus proteins and host cell membranes during the viral life cycle.* Int Rev Cytol, 2005. **245**: p. 171-244.
79. Thomas, P.D., et al., *PANTHER: Making genome-scale phylogenetics accessible to all.* Protein Sci, 2022. **31**(1): p. 8-22.
80. Michalski, D., et al., *Zika virus noncoding sfRNAs sequester multiple host-derived RNA-binding proteins and modulate mRNA decay and splicing during infection.* J Biol Chem, 2019. **294**(44): p. 16282-16296.
81. Soto-Acosta, R., et al., *Fragile X mental retardation protein is a Zika virus restriction factor that is antagonized by subgenomic flaviviral RNA.* Elife, 2018. **7**.
82. Wang, Y., et al., *RNA Helicase A Is an Important Host Factor Involved in Dengue Virus Replication.* Journal of Virology, 2019. **93**(4): p. 10.1128/jvi.01306-18.
83. Szklarczyk, D., et al., *STRING v10: protein-protein interaction networks, integrated over the tree of life.* Nucleic Acids Res, 2015. **43**(Database issue): p. D447-52.
84. Nourreddine, S., et al., *NF45 and NF90 Regulate Mitotic Gene Expression by Competing with Staufen-Mediated mRNA Decay.* Cell Reports, 2020. **31**(7): p. 107660.
85. Graber, T.E., et al., *NF45 functions as an IRES trans-acting factor that is required for translation of cIAP1 during the unfolded protein response.* Cell Death & Differentiation, 2010. **17**(4): p. 719-729.
86. Schmidt, T., et al., *NF90-NF45 is a selective RNA chaperone that rearranges viral and cellular riboswitches: biochemical analysis of a virus host factor activity.* Nucleic Acids Res, 2017. **45**(21): p. 12441-12454.
87. Reichman, T.W., L.C. Muñiz, and M.B. Mathews, *The RNA binding protein nuclear factor 90 functions as both a positive and negative regulator of gene expression in mammalian cells.* Mol Cell Biol, 2002. **22**(1): p. 343-56.
88. Isken, O., et al., *Nuclear factors are involved in hepatitis C virus RNA replication.* Rna, 2007. **13**(10): p. 1675-92.
89. Isken, O., et al., *Members of the NF90/NFAR protein group are involved in the life cycle of a positive-strand RNA virus.* Embo j, 2003. **22**(21): p. 5655-65.
90. Gomila, R.C., G.W. Martin, III, and L. Gehrke, *NF90 Binds the Dengue Virus RNA 3' Terminus and Is a Positive Regulator of Dengue Virus Replication.* PLOS ONE, 2011. **6**(2): p. e16687.
91. Isken, O., et al., *Complex signals in the genomic 3' nontranslated region of bovine viral diarrhea virus coordinate translation and replication of the viral RNA.* Rna, 2004. **10**(10): p. 1637-52.
92. Wolkowicz, U.M. and A.G. Cook, *NF45 dimerizes with NF90, Zfr and SPNR via a conserved domain that has a nucleotidyltransferase fold.* Nucleic Acids Research, 2012. **40**(18): p. 9356-9368.
93. Van Bortle, K., et al., *A cancer-associated RNA polymerase III identity drives robust transcription and expression of snaR-A noncoding RNA.* Nature Communications, 2022. **13**(1): p. 3007.
94. Castella, S., et al., *Ilf3 and NF90 functions in RNA biology.* WIREs RNA, 2015. **6**(2): p. 243-256.
95. Parrott, A.M., et al., *The evolution and expression of the snaR family of small non-coding RNAs.* Nucleic Acids Research, 2010. **39**(4): p. 1485-1500.

96. Tonello, F., M.L. Massimino, and C. Peggion, *Nucleolin: a cell portal for viruses, bacteria, and toxins.* Cellular and Molecular Life Sciences, 2022. **79**(5): p. 271.
97. Ma, N., et al., *Nucleolin functions in nucleolus formation and chromosome congression.* Journal of Cell Science, 2007. **120**(12): p. 2091-2105.
98. Ginisty, H., et al., *Structure and functions of nucleolin.* Journal of Cell Science, 1999. **112**(6): p. 761-772.
99. Tajrishi, M.M., R. Tuteja, and N. Tuteja, *Nucleolin: The most abundant multifunctional phosphoprotein of nucleolus.* Commun Integr Biol, 2011. **4**(3): p. 267-75.
100. Ginisty, H., et al., *Structure and functions of nucleolin.* J Cell Sci, 1999. **112 (Pt 6)**: p. 761-72.
101. Balinsky, C.A., et al., *Nucleolin interacts with the dengue virus capsid protein and plays a role in formation of infectious virus particles.* J Virol, 2013. **87**(24): p. 13094-106.
102. Bian, W.-X., et al., *Binding of cellular nucleolin with the viral core RNA G-quadruplex structure suppresses HCV replication.* Nucleic Acids Research, 2018. **47**(1): p. 56-68.
103. Hajikhezri, Z., et al., *Role of CCCH-Type Zinc Finger Proteins in Human Adenovirus Infections.* Viruses, 2020. **12**(11).
104. Zhang, B., et al., *Zinc Finger CCCH-Type Antiviral Protein 1 Restricts the Viral Replication by Positively Regulating Type I Interferon Response.* Frontiers in Microbiology, 2020. **11**.
105. Zhu, M., et al., *The CCCH-Type Zinc Finger Antiviral Protein Relieves Immunosuppression of T Cells Induced by Avian Leukosis Virus Subgroup J via the NLP-PKC-δ-NFAT Pathway.* J Virol, 2022. **96**(2): p. e0134421.
106. Zhu, M., et al., *CCCH-type zinc finger antiviral protein mediates antiviral immune response by activating T cells.* Journal of Leukocyte Biology, 2020. **107**(2): p. 299-307.
107. Plant, E.P. and Z. Ye, *Bias at the third nucleotide of codon pairs in virus and host genomes.* Sci Rep, 2022. **12**(1): p. 4522.
108. Udenze, D., et al., *CpG content in the Zika virus genome affects infection phenotypes in the adult brain and fetal lymph nodes.* Frontiers in Immunology, 2022. **13**.
109. Sauter, D. and F. Kirchhoff, *Less is more: Biased loss of CpG dinucleotides strengthens antiviral immunity.* PLOS Biology, 2021. **19**(9): p. e3001353.
110. Majumder, M., R.H. Johnson, and V. Palanisamy, *Fragile X-related protein family: a double-edged sword in neurodevelopmental disorders and cancer.* Crit Rev Biochem Mol Biol, 2020. **55**(5): p. 409-424.
111. Chen, E. and S. Joseph, *Fragile X mental retardation protein: A paradigm for translational control by RNA-binding proteins.* Biochimie, 2015. **114**: p. 147-54.
112. Cave, J.W. and D.E. Willis, *G-quadruplex regulation of neural gene expression.* The FEBS Journal, 2022. **289**(12): p. 3284-3303.
113. Ceman, S., V. Brown, and S.T. Warren, *Isolation of an FMRP-associated messenger ribonucleoprotein particle and identification of nucleolin and the fragile X-related proteins as components of the complex.* Mol Cell Biol, 1999. **19**(12): p. 7925-32.
114. Darnell, J.C., et al., *Fragile X Mental Retardation Protein Targets G Quartet mRNAs Important for Neuronal Function.* Cell, 2001. **107**(4): p. 489-499.
115. Richter, J.D. and X. Zhao, *The molecular biology of FMRP: new insights into fragile X syndrome.* Nat Rev Neurosci, 2021. **22**(4): p. 209-222.

116. Tsang, B., et al., *Phosphoregulated FMRP phase separation models activity-dependent translation through bidirectional control of mRNA granule formation.* Proceedings of the National Academy of Sciences, 2019. **116**(10): p. 4218-4227.
117. Garcia-Jove Navarro, M., et al., *RNA is a critical element for the sizing and the composition of phase-separated RNA–protein condensates.* Nature Communications, 2019. **10**(1): p. 3230.
118. Guo, Q., X. Shi, and X. Wang, *RNA and liquid-liquid phase separation.* Non-coding RNA Research, 2021. **6**(2): p. 92-99.
119. Joseph, J.A., et al., *Thermodynamics and kinetics of phase separation of protein-RNA mixtures by a minimal model.* Biophys J, 2021. **120**(7): p. 1219-1230.
120. Van Lindt, J., et al., *A generic approach to study the kinetics of liquid–liquid phase separation under near-native conditions.* Communications Biology, 2021. **4**(1): p. 77.
121. He, C., et al., *DDX17 modulates the expression and alternative splicing of genes involved in apoptosis and proliferation in lung adenocarcinoma cells.* PeerJ, 2022. **10**: p. e13895.
122. Boleslavska, B., et al., *DDX17 helicase promotes resolution of R-loop-mediated transcription-replication conflicts in human cells.* Nucleic Acids Res, 2022. **50**(21): p. 12274-12290.
123. Giraud, G., S. Terrone, and C.F. Bourgeois, *Functions of DEAD box RNA helicases DDX5 and DDX17 in chromatin organization and transcriptional regulation.* BMB Rep, 2018. **51**(12): p. 613-622.
124. Ngo, T.D., A.C. Partin, and Y. Nam, *RNA Specificity and Autoregulation of DDX17, a Modulator of MicroRNA Biogenesis.* Cell Rep, 2019. **29**(12): p. 4024-4035.e5.
125. Lambert, M.-P., et al., *The RNA helicase DDX17 controls the transcriptional activity of REST and the expression of proneural microRNAs in neuronal differentiation.* Nucleic Acids Research, 2018. **46**(15): p. 7686-7700.
126. Sourabh, S., M. Chauhan, and R. Tuteja, *Chapter 6 - Genome Wide In Silico Characterization of Ded1 Family of Helicases from Plasmodium Falciparum*, in *Helicases from All Domains of Life*, R. Tuteja, Editor. 2019, Academic Press. p. 97-112.

Chapter 5. Nucleolin interacts with the 5’ and 3’ terminal regions of the Zika virus and is critical for the formation of functional virions ***Contribution percentage = 85%***

Danielle Gemmill[1], Higor Pereira[1], Clément Mazeaud[2], Quadir Siddiqui[1], Laurent Chatel-Chaix[2], and Trushar Patel[1,3,4]

[1]Alberta RNA Research and Training Institute & Department of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, Alberta.
[2]Institut National de la Recherche Scientifique, Centre INRS-Institut Armand-Frappier, Laval, Quebec, Canada.
[3]Department of Microbiology, Immunology and Infectious Disease, Cumming School of Medicine, University of Calgary, Alberta, Canada.
[4]Li Ka Shing Institute of Virology and Discovery Lab, University of Alberta, Edmonton, Alberta

5.1 Abstract

Zika virus (ZIKV) is a neurovirulent arbovirus that can cause several neurological sequelae, including gestational microencephaly through vertical transmission of a pregnant individual. The 5’ and 3’ untranslated regions of the genome, known as the terminal regions, are highly conserved and critically involved in the regulation of transcription and translation of the viral genome, as well as suppression of host pathways to further enhance viral amplification. Currently, there are no treatments or vaccines available for ZIKV, and novel treatments must be explored to combat the public health burden ZIKV causes. To investigate this, we performed an immunoprecipitation RNA-baited pulldown assay to identify the human neuronal proteins that interact with the terminal regions to understand the direct binding partners that are required for the virus to replicate, altogether hoping to provide novel targets for ZIKV to stifle replication. We identified Nucleolin (NCL) as one of the binding partners and we biophysically characterized its structure as well as the ZIKV terminal regions by performing multi-angle/dynamic light scattering, circular dichroism spectroscopy, and small-angle X-ray scattering. We then characterized the interaction of NCL and the terminal regions using electrophoretic mobility shift assays, and

microscale thermophoresis. Finally, in collaboration with the Chatel-Chaix research group at INRS, Quebec, we performed a siRNA knockdown of NCL in the Huh-7.5 hepatocarcinoma cell line and infected the cells with ZIKV to understand if the virus can replicate in a NCL-free environment. We show for the first time that NCL is critical for the formation of functional virions, further providing a potential target to stop ZIKV replication, thus, easing the burden on public health.

5.2 Introduction

ZIKV is a mosquito-borne flavivirus that horizontally transmits to humans when mosquitos are extracting blood for a meal [1]. The virus disperses from the saliva of the mosquito into the host's bloodstream and interacts with susceptible cells containing tyrosine kinase (TAM) receptors [2]. The virus then replicates inside permissive cells containing the correct biomolecular landscape that favours the replication [3]. Through viral tropistic investigation using immunohistochemistry, virions have been found concentrated in neuronal tissues and have been shown to cause a host of neurological sequelae in all age groups [4]. Moreover, this was alarmingly noticed in developing fetuses during the 2015/2016 outbreak of ZIKV in South America, wherein gestational microencephaly cases significantly increased during the outbreak, sparking an investigation into the sudden spike [5]. Researchers discovered that the virus was vertically transmitted from the pregnant individual to the developing fetus, causing severe disruptions to fetal neuronal development [6, 7]. Research has gone into the development of a vaccine, but unfortunately, there are significant concerns about the cross-reactivity of antibodies with Dengue virus (DENV) expanding the tropism of DENV and causing severe illness and/or death; a phenomenon known as antibody-dependent enhancement (ADE) of infection [8-11]. This event prevents the development of more traditional vaccines, and more investigation into newly developed vaccine approaches, such as mRNA vaccines are being investigated [12]. In tandem with this, antiviral treatments are also being investigated to slow viral replication during infection to allow the immune system to clear the virus more rapidly [13-15].

ZIKV contains a positive-sense RNA genome approximately ~10.8 kilobases long, flanked with highly conserved, cis-acting, non-coding terminal regions (TRs) [16]. These specific regions have been shown to completely regulate the entire fate of the viral genome in the cytoplasm by controlling the transcription and translation of the virus [17-19]. This is performed by the hijacking of host proteins, coupled with base pairing to each other by a process known as genomic cyclization [17-19]. This molecular switch inhibits/allows initiation of transcription or translation of the viral genome [17-19]. There have been efforts to elucidate the functions of the flaviviral TRs in general, as they remain the most conserved region in the entire flaviviral family, which implies evolutionary criticality for survival, functioning, and/or adaptation to host-switching [18, 20, 21]. With the TRs being conserved, this allows for an exploitative target that can be used by researchers to investigate novel antiviral therapies. Moreover, targeting a viral family with these consensus sequences/structures could potentially be a multi-species treatment for flavivirus, thus lifting a public health and socioeconomic burden.

Recently, there have been investigations into antiviral therapies that target host protein-viral RNA interfaces, as seen in recent studies of antiviral targets in SARS-CoV-2, Hepatitis C virus, Influenza virus, Herpesvirus, and DENV [22-26]. In the SARS-CoV-2 investigation, the overall process allowed for the repurposing of 59 compounds already approved by the FDA that had an impact on viral replication [24, 26]. Moreover, in tandem with exploring previously developed drugs to target viral replication, expanding the boundaries of therapeutics must be additionally pursued, as many viruses still do not have treatment options, ZIKV being one of them [27].

We aimed to investigate the human proteins that interact with the 5’ and 3’ TRs of ZIKV, allowing for more knowledge to explore existing or novel therapeutics to target the virus, nullifying the issue of ADE, and directly slowing down the virus from replicating and allowing the immune response to clear viral particles. To directly identify the human protein binding partners interacting specifically with the TRs, coupled with tropistic relevance to neuronal tissue, we performed an immunoprecipitation ‘RNA-baited’ pulldown assay using the neuronal cell line, SH-SY5Y, as well as using the ZIVK 5’ and 3’ TRs as bait (see chapter 4). One of the protein binding partners we identified was NCL. However, the identification of a protein via pulldown assay could also indicate that it is indirectly interacting through protein-protein interactions [28].

NCL is a eukaryotic multifunctional phosphoprotein involved in a myriad of cellular functions, such as (but not limited to) intrinsic DNA and RNA helicase activity, self-cleaving activities, binding to DNA and RNA through its RNA recognition motifs (RRMs), assisting in regulating DNA and RNA metabolism, chromatin structure, rDNA transcription, rRNA maturation, cytokinesis, nucleogenesis, cell proliferation and growth, as well as the folding, maturation and ribosome assembly, and finally, nucleocytoplasmic transport of newly synthesized pre-RNAs [29, 30]. Conversely, NCL is also hijacked by various viruses and bacteria during host infection, as it’s been shown to be exploited by Influenza A, Respiratory Syncytial, Human Immunodeficiency, and Hepatitis C viruses, and *Pseudomonas aeruginosa*, *Escherichia coli*, and *Staphylococcus aureus* bacteria [31-34]. With the identification of NCL in the pulldown assays, as well as previous literature indicating that NCL is exploited by RNA viruses in general, we aim to explore and characterize the interactions that NCL make with the highly conserved TRs of ZIKV, and as well, elucidate if NCL is critical for ZIKV virion formation.

We investigated if $NCL_{290-652}$ (truncated with the protein sequence between these amino acids) is a direct binding partner to the 5’ and 3’ TRs, followed by a characterization of the structure and kinetics of said interaction. To structurally characterize $NCL_{290-652}$, we performed circular dichroism spectroscopy (CD spec) to gain insight into the secondary structure profile and to validate if it is functionally folded. We also performed multi-angle/dynamic light scattering (MALS/DLS) and small-angle X-ray scattering (SAXS) to retrieve data on $NCL_{290-652}$ and the TR’s monodispersity and general shape and size in solution. We then implemented an RNA tiling approach to understand if there were kinetic differences within specific regions of the 3’ TR, specifically, the 3’ xrRNA 1 & 2, 3’ dumbbells, and 3’ stem-loop (Fig. 5.1). Finally, we investigated if NCL is critical for viral replication by performing gene knockdowns of NCL in the human Huh-7.5 cells along with cell-based virology assays with ZIKV and DENV in collaboration with the Chatel-Chaix research group at the University of Laval. We measured plaque formation, as well as quantification of viral RNA to understand if the number of functional virions as well as viral RNA are collectively decreased.

5.3 Materials and Methods

5.3.1 ZIKV 5’ and 3’ TRs and $NCL_{290-652}$ expression and purification to homogeneity

The ZIKV 5’ and 3’ TR transcript sequences were extracted from a ZIKV isolate from Haiti (GenBank: KU509998.3), then cloned into pUC57-Kan plasmids constructed at Integrated DNA

Technologies™ designed with a T7 polymerase promotor upstream and an *XbaI* restriction enzyme cut site downstream of the transcript. This allowed for higher *in vitro* transcription (IVT) yields with a T7 polymerase providing termination immediately after *XbaI* digestion incubation, altogether creating a linearized plasmid to allow T7 to dissociate from the DNA at the end of the transcript. The plasmids were transformed into NEB5α *E. coli* cells to amplify the plasmid amount to provide a high yield for IVTs. The IVTs were performed using IVT buffer (described in previous literature) and incubated for 3 hours at 37°C [38]. The IVT reaction products were then purified using a Superdex® 200 Increase 10/300 GL for the 5' TR, 3' xrRNA 1 & 2, 3' dumbbells, and 3' stem-loop, and a Sephacryl® S-400 HR for the 3' TR size exclusion chromatography (SEC) columns both from Cytiva©, which were altogether performed to purify the transcripts from other IVT components (Fig. 5.1). A 2% agarose gel was run to check for pure fractions, which were then pooled, vitrified, and stored at -80°C until use.

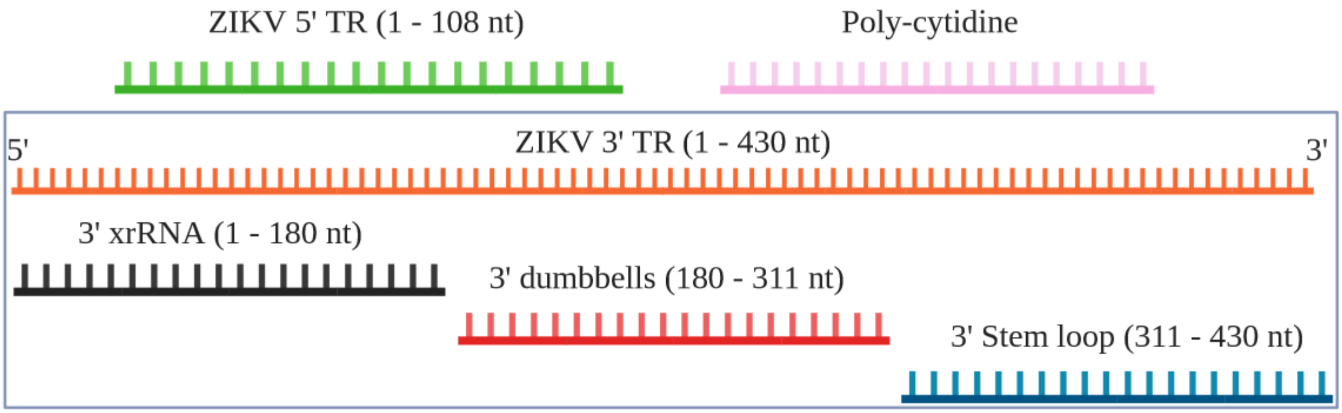


Figure 5.1. A schematic of the 5' (green) and 3' (orange) TR transcripts of ZIKV, along with the RNA tiled transcripts of the 3' xrRNA 1 & 2 (black), 3' dumbbells (red), and 3' stem-loop (blue). Additionally, the poly-cytidine (pink) is present for the negative control (created with BioRender.com).

The $NCL_{290-652}$ construct was commercially synthesized from Genewiz®-Aszenta Life Sciences® using a pET28-a vector, and designed between amino acids 290-652, constituting the 4 RRMs (Fig. 5.2). Additionally, we accommodated slight overlaps of disordered regions on either side of the RRMs to allow overall flexibility in the domains/structure, as well as provide an effective C-terminal hexahistidine tag for the immobilized metal affinity chromatography (IMAC) resin. The large, disordered regions of NCL are from amino acids 1-299 and 640-710, which were altogether removed to streamline protein overexpression in the *E. coli* BL21 (DE3) cells containing the Lemo System™ (Fig. 5.2). The transformed cells were grown at 37°C until they reached OD_{600} followed by induction of 1 mM isopropylthio-β-galactoside (IPTG). The temperature was then decreased to 16°C and incubated for 16 hours, followed by centrifugal harvest of the cells. The cells were lysed using Phosphate Buffer Saline (PBS) along with 10 mM imidazole, 0.1 mg/mL lysozyme, 5,200 units of DNase I (Thermofisher™), 2 mM PMSF, as well as other protease inhibitors (pepstatin, aprotinin, and AEBSF) from Biobasic™. Following lysis, sonication was performed with 10 sec on 15 off, 60% Amplitude, with 25 cycles. Clarification was then done at 30,000 x g, and 0.45 µm filtration was performed prior to loading the lysate onto the IMAC Ni^{2+} column (1 mL Histrap™ high-performance column from Cytiva™ mounted on an Äkta Start™ system). The sample was purified using a protocol of 2 washes at 5 and 10 mM imidazole at 10 column volumes, respectively, followed by an elution gradient from 10 mM to 150 mM imidazole supplemented with the lysis buffer over a total of 25 column volumes (25 mL). The fractions containing $NCL_{290-652}$ were run on a 12% SDS-PAGE and those that contained $NCL_{290-652}$ were collected from the IMAC fractions and run through a MonoQ™ anion exchange column to extract nucleic acid contaminants by charge-charge interactions with the resin. Confirmed PAGE fractions were then followed through with a Superdex™ 75 increase 10/300 GL (Cytiva™) SEC column pre-equilibrated with 1x PBS mounted on an Äkta Pure™ system. The purified samples were then concentrated to 30 µM, vitrified, and stored at -80°C until further use.

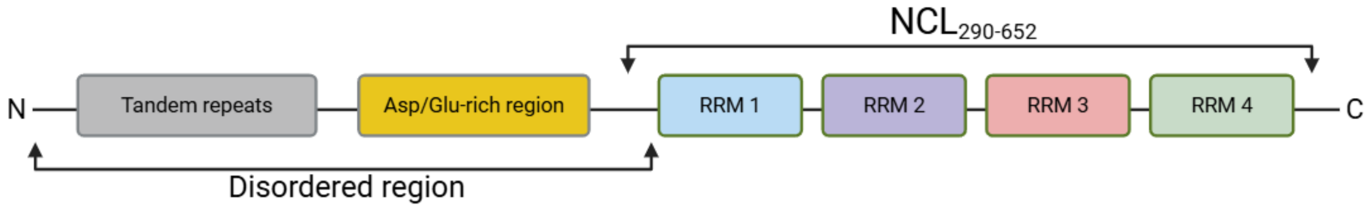


Figure 5.2. A schematic representation of the various domains of NCL, along with the selected area that we utilized in our studies collectively excludes the remainder of the protein which has been predicted to be intrinsically disordered [39-41] (created with BioRender.com).

5.3.2 CD spectroscopy of NCL$_{290-652}$

A Jasco™ J-815 spectropolarimeter (Jasco Inc™, Easton, MD) was used to collect spectra of NCL$_{290-652}$ ranging from 220 to 320 nm, using a 1.0 mm cell, 0.1 nm data pitch, with five accumulations and 32 s integration time. All measurements were baseline corrected with degassed 1x PBS and repeated in triplicate at 25°C. The protein sample in 1x PBS was prepared by concentrating the species to 40 µM. The raw data was collected using the Jasco™ Spectra Manager software and the dataset was then exported to the K2D2 software for deconvolution to reveal the percentage of secondary structure of NCL$_{290-652}$ [42]. The deconvoluted secondary structures were compared against the submitted European Molecular Biology Laboratory - European Bioinformatics Institute (EMBL-EBI) Protein Data Bank structure of human NCL$_{290-652}$, and the percentage of secondary structure was compared [43-47] (Table 5.1).

5.3.3 Multi-angle/dynamic light scattering of NCL$_{290-652}$

Purified NCL$_{290-652}$ was loaded onto a 1x PBS equilibrated Shodex™ KW-404F (Resonac America, Inc.©) using a Vanquish HPLC from ThermoFisher Scientific™. The Shodex™ column was placed in series with a Dawn© (Wyatt Technology Corporation™) multi-angle light scattering (MALS) instrument with the laser source producing a 658 nm wavelength for particle scattering. An Optilab™ dynamic light scattering (DLS) (Wyatt Technology by Waters Corporation™) refractometer was used to measure the solvent refractive index and absolute concentration of solutes at 658 nm wavelength using a refractive index increment $\left(\frac{dn}{dc}\right)$ of 0.185 mL/g for protein measurement [48, 49]. The data analysis was performed using the ASTRA™ v8.0.0.25 program, which calculated the absolute molecular weight (MW) of NCL$_{290\text{-}652}$ using the equation:

$$Mw = \frac{R(\theta)}{KC\left(\frac{dn}{dc}\right)^2}$$

Where for each elution point taken by the software, R(θ) is Rayleigh's ratio, K is the polymer constant, C is the weight concentration of the solution, and $\left(\frac{dn}{dc}\right)$ is the refractive index increment.

5.3.4 Small-angle X-ray scattering of the ZIKV TRs and NCL$_{290\text{-}652}$

SAXS was performed using the B21 HPLC-SAXS beamline at Diamond Light Source (Didcot, Oxfordshire, UK), as reported elsewhere [50]. An Agilent 1200 (Agilent Technologies, Stockport, UK) HPLC was utilized through connection to a specialized flow cell, and 50 µL of each NCL$_{290\text{-}652}$ and ZIKV 5' and 3' TR were individually injected into a 1x PBS buffer for NCL$_{290\text{-}652}$ and 10 mM bis-tris pH 7, 100 mM NaCl, 15 mM KCl, and 15 mM MgCl$_2$ for the ZIKV RNA, equilibrated Shodex 403KW-4F HPLC column (Showa Denko America Inc., New York, NY, USA) with a flow rate of 0.160 mL/min. 600 frames with 3-sec exposure to X-rays were collected for both the TRs and NCL$_{290\text{-}652}$. The scattering data was analyzed using the ATSAS suite [51]. The collected data for the TRs and

$NCL_{290-652}$ were buffer subtracted by taking the sample peak in CHROMatography Inline X-ray Scattering (CHROMIXS) [52]. Following this, analysis of the sample peak was performed using the ATSAS suite [53]. A quality check for the TRs and $NCL_{290-652}$ as well as the radius of gyration (R_g) were assessed by performing a Guinier analysis [54]. A dimensionless Kratky analysis was then performed which showed the general globularity and linear profile of the TRs and $NCL_{290-652}$, followed by a paired-distribution function (P(r)) to provide real-space R_g and the maximum particle dimension (D_{max}) using GNOM [55-60]. The P(r) plot was used as the input data for the DAMMIN [61] software, which calculated and presented predicted structures, followed by the DAMAVER [61] software for obtaining an average model of the predicted structures. Finally, the DAMFILT [62] software yielded filtered representative structures of both the TRs and $NCL_{290-652}$ (Fig. 5.8).

5.3.5 Electrophoretic Mobility Shift assays, and Microscale thermophoresis kinetic studies

The EMSAs were run using a constant concentration of ZIKV TRs at 500 nM, and 8x 2-fold serial dilutions of $NCL_{290-652}$ starting at 30 µM. The RNA and protein were incubated at ambient room temperature for 30 minutes prior to loading onto the 6% native PAGE (30% 29:1 acrylamide:bisacrylamide ratio in 50 mM potassium phosphate (pH 7)). The gels were run using 50mM potassium phosphate (pH 7) running buffer and run for 45 min and 1.5 hours for the 5' and 3' TRs respectively, and together run at 150 volts. Band visualization was performed by staining with SYBR-safe nucleic acid dye (ThermoFisher™) after the gel was run to avoid disruption to the interaction.

All MST experiments were performed on a NanoTemper Technologies™ Monolith NT.115 instrument at 22°C with the labelled RNA concentrations constant at 150 nM, and $NCL_{290-652}$ concentration 6 µM – 732 pM, altogether serially diluted 2-fold across 14 tubes (n=3) [63]. Prior to data collection, the RNA-protein samples were mixed and incubated at room temperature for 30 min. The excitation power was set to 100% with medium infrared laser power (60%). The collected data was then inputted into the MO.Affinity Analysis v2.3 Software provided by NanoTemper Technologies™ and analyzed for the determination of the K_d. Finally, a poly-cytidine oligo was used as a negative control, as it's been shown to be an effective negative control in NCL interaction studies [64, 65].

5.3.6 Human cell-based assays - Chatel-Chaix group

The Chatel-Chaix research group performed all the ZIKV cell-based assays on human hepatocarcinoma Huh-7.5 cells using the H/PF/2013 ZIKV isolate and the DENV2 NGC strain. The Huh 7.5 cells were cultured and supplemented with DMEM (ThermoFisher™) supplemented with 10% fetal bovine serum (Wisent™), 1% non-essential amino acids (ThermoFisher™) and 1% penicillin-streptomycin (ThermoFisher™) [66]. The ZIKV and DENV NCL knockdowns were performed using siRNA transfected into the Huh-7.5 cells, followed by western blotting using anti-NCL antibodies. This process was repeated as well for ZIKV and DENV-infected cells. The plaque assays were performed one day after infection, and the samples were serially diluted to 10^{-1} at 10^{-6}-fold in complete DMEM [66]. The cells were fixed for 2 hours in 2.5% formaldehyde, following by staining with crystal violet/10% ethanol for 15 min. The newly stained cells were then subjected to plaque counting, and the titers of virions were reported in PFU/mL [66]. The

DENV and ZIKV viral RNA transcript quantification was performed using RT-pPCR with the total RNA being extracted from cells using the RNeasy Mini Kit (Qiagen) [66]. The RNA was then subjected to reverse transcription using Invitrogen SuperScript™ IV, and real-time PCR was performed using the Applied Biosystems SYBR Green Master mix (Life Technologies) and a LightCycler 96 (Roche) for detection [66].

5.4 Results and Discussion

5.4.1 ZIKV 5’ and 3’ TR, and $NCL_{290-652}$ purification

The ZIKV TRs and 3’ TR sections were *in vitro* transcribed according to previously published data from our lab [20, 67-69]. All RNAs were subjected to SEC (Fig. 5.3). The ZIKV 3’ SL product is present in our recent publication, as the identical RNA batch was used [38]. Furthermore, the elution profiles vary for each RNA, with the 5’ TR eluting at ~13 mL, the 3’ xrRNA at ~13mL, the 3’ DB at ~15mL, and the 3’ SL at ~13.5 mL altogether run on the Superdex™ 200 (Fig. 5.3). The elution peak of the 3’ TR on the Sephacryl™ 400 was shown to be at ~16 mL (Fig. 5.3). Moreover, all RNAs were run on a 2% agarose gel to validate size, as well as oligomeric state (Fig. 5.3). Only fractions containing monomers and at the correct size were collected for downstream experiments. After collection and vitrification of each RNA, they were subjected to labelling reactions with Alexa™ 488, followed by an additional SEC purification via a Superdex™ 200 to separate unbound dye from the RNA as well as separation of monomeric RNA species.

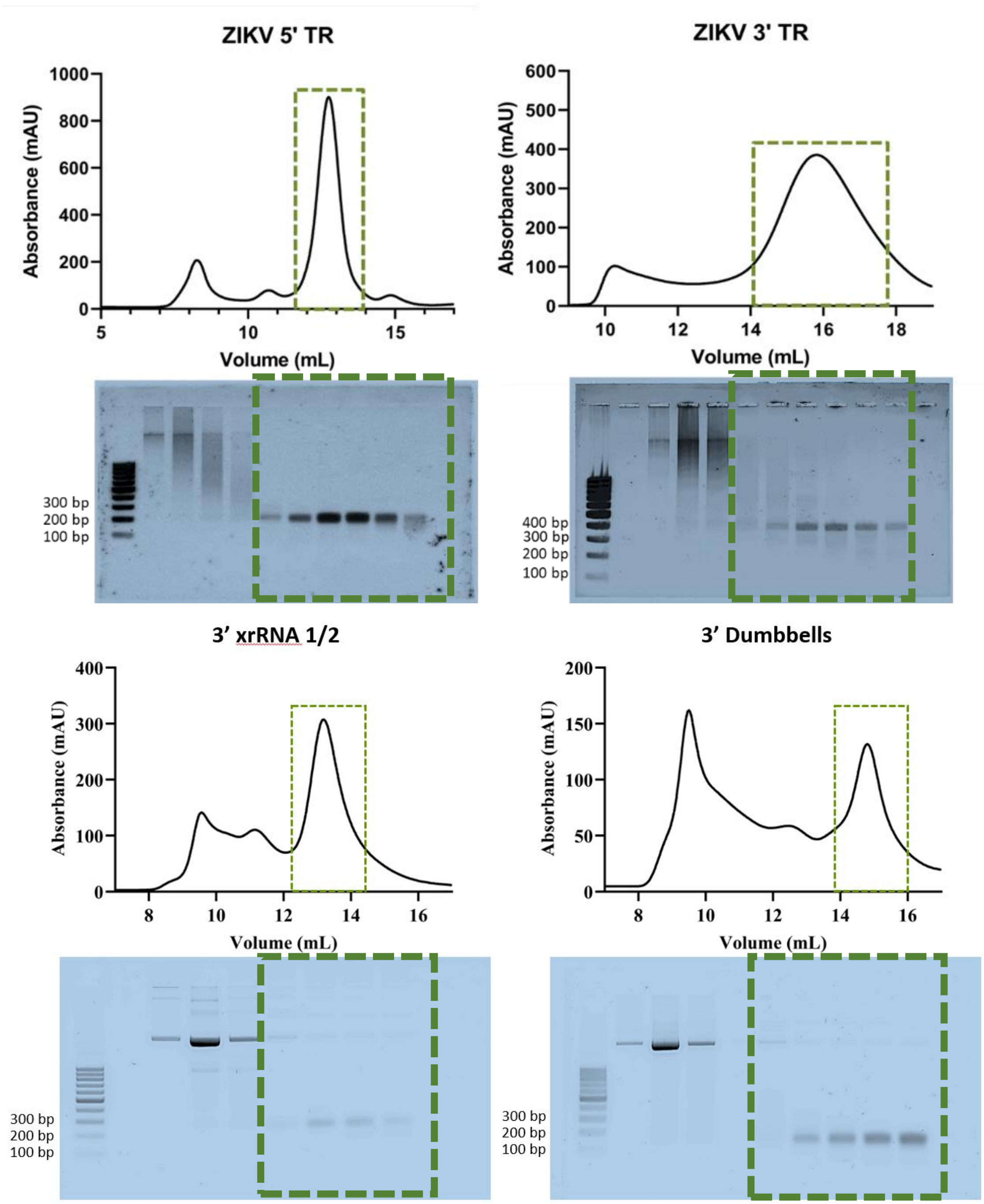


Figure 5.3. SEC chromatograms of the 5’ and 3’ TRs as well as the 3’ xrRNA 1 & 2, and 3’ dumbbells. The 3’ SL was transcribed and purified in previous literature [38]. The green boxes indicate fractions that were taken for downstream experimentation that contain RNA, which were fractions only containing the correct size of transcript and where a single band were taken for further experiments. The fractions were run on 2% agarose gels (below each respective chromatogram) and run next to the plasmid elution profile to indicate the size difference between the species.

The overexpression of $NCL_{290\text{-}652}$ in Lemo™ 21 *E. coli* cells yielded ~15 grams of cells from 1L of LB culture, together incubated for 16 hours in 1mM IPTG. The lysed, clarified, and filtered product was loaded onto the 5 mL IMAC column, allowing the C-terminal his-tag to interact with the nickel (Fig. 5.4). The stepwise elution with 20 and 300 mM imidazole allowed for loosely bound species to be separated from $NCL_{290\text{-}652}$ (Fig. 5.4). The protein eluted at 300mM imidazole. The selected area was then run on a 12% SDS-PAGE to determine the fractions containing high yields of $NCL_{290\text{-}652}$ which were then pooled and immediately loaded onto the MonoQ™ anion exchange column wherein any nucleic acid contaminants were extracted from the sample by charge-charge interactions with the resin (Fig. 5.4). The substantial difference in overall negative charge of $NCL_{290\text{-}652}$ (pI = 5.07) and the nucleic acids allowed separation and further purity to the overall $NCL_{290\text{-}652}$ sample (Fig. 5.4). The $NCL_{290\text{-}652}$ sample eluted from the MonoQ™ at ~47 mL at ~35% elution buffer B (PBS with ~350 mM NaCl). Following anion exchange, Further purification was performed via SEC to ensure all contaminated species were separated from $NCL_{290\text{-}652}$ monomers, and any partially degraded product, and provide buffer exchange to 1x PBS. Monomeric $NCL_{290\text{-}652}$ was shown to elute at ~14 mL on the Superdex™ 75, confirmed via 12% SDS-PAGE (Fig. 5.4). Following this, the pure $NCL_{290\text{-}652}$ was concentrated and vitrified for storage until downstream experiments were performed.

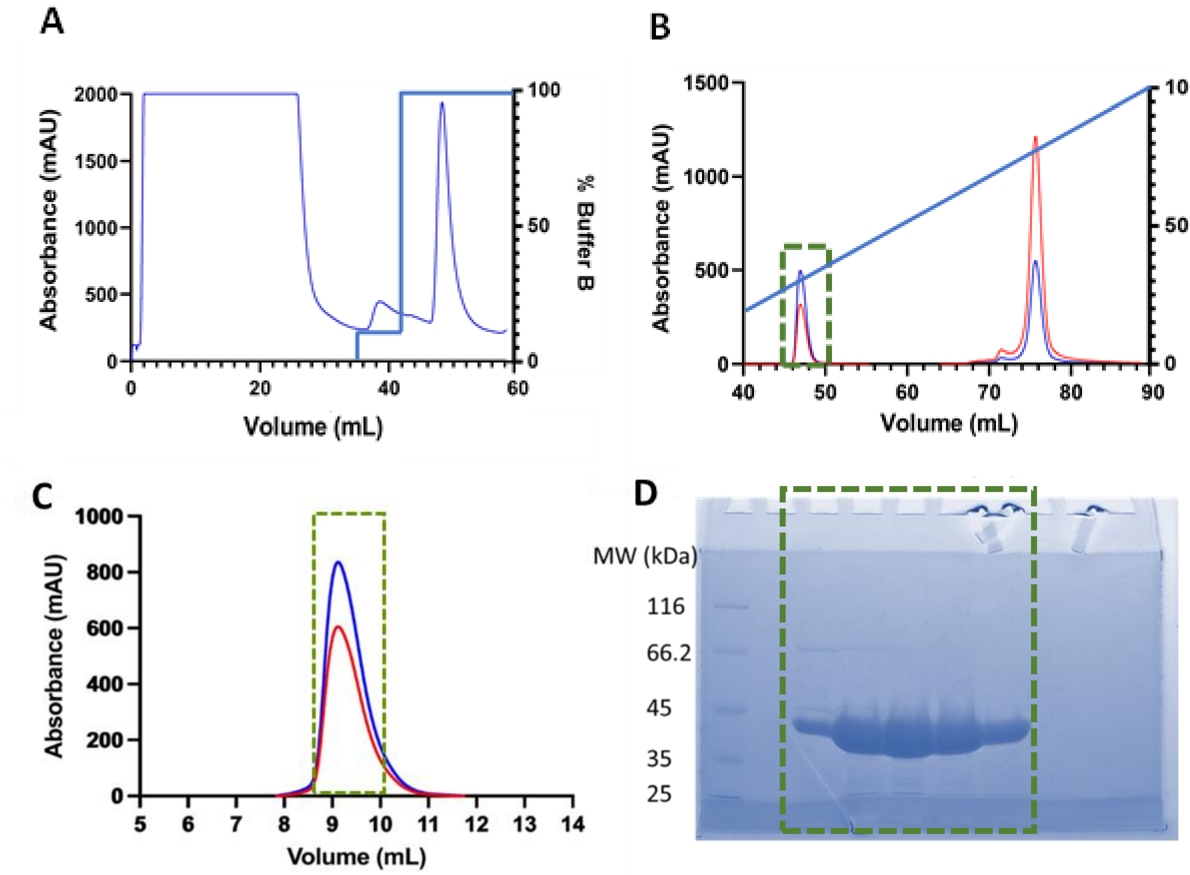


Figure 5.4. The purification process of NCL$_{290-652}$. Panels **A**, **B**, and **C** are IMAC, anion exchange, and SEC chromatograms, respectively. The green boxes indicated in Panels **A**, **B**, and **C** indicate the regions that were taken for downstream purification, and then storage for downstream experiments. Panel **A** is the IMAC stepwise elution (20mM & 300mM of imidazole) of NCL$_{290-652}$, and panel **B** shows the absorbances at 260 (red) and 280 nm (blue) while a gradient of 0-500mM NaCl is applied. Panel **C** is the SEC purification alongside with Panel **D**, its corresponding 12% SDS-PAGE run after SEC of NCL$_{290-652}$, with the molecular weight marker beside to indicate the molecular weight. The prominent bands are hypothesized to be NCL$_{290-652}$, which were collectively taken for further experimentation. The green squares altogether indicate the fraction area taken for downstream processes.

5.4.2 Structural investigation reveals NCL$_{290-652}$ forms a monomer in solution

NCL$_{290-652}$ was measured in the MALS and DLS instruments which generally allows researchers to gain insight into the physical properties of biomolecules in solution species, such as the oligomeric state, relative size (R$_h$), and the absolute molecular weight (kDa) [73]. Using MALS/DLS coupled to an HPLC to produce flow through the instruments for data collection, as well as provide SEC purification, we determined that $NCL_{290-652}$ predominantly adopts a monomeric state in solution. This is shown by the calculated MW from the selected scattering peak which was determined to be ~46.88 kDa and the theoretical $NCL_{290-652}$ being ~42.3 kDa (Fig. 5.5). This is an acceptable range, which is a margin of error of <± 10%, which is an optimal error when using SEC coupled with MALS, factoring in potential column interactions and/or data analysis [74, 75]. To clarify if this slight addition in MW was due to any contamination, light spectroscopy and a 2% agarose gel were run to ensure the $NCL_{290-652}$ sample was free of nucleic acid species, which was discovered to be free of detectable contamination (Fig. 5.8). Furthermore, the species was passed through the DLS to determine the radius of hydration (R_h) based on the light scattering fluctuations over time as the sample experiences Brownian motion as it passes through the detection source [76]. Based on the data collected, $NCL_{290-652}$ was determined to have an R_h of 3.56 nm Figure 5.6. The ZIKV 5' and 3' TRs still require optimization for SEC-MALS, and the TRs and $NCL_{290-652}$ mixed for data collection at high concentrations have had a difficult time providing reliable data, as the samples are experiencing oligomerization in the form of aggregates or, potentially an RNA phase separation event, as is commonly seen with NCL when it is interacting with various nucleic acids [35-37]. More investigation and optimization are needed.

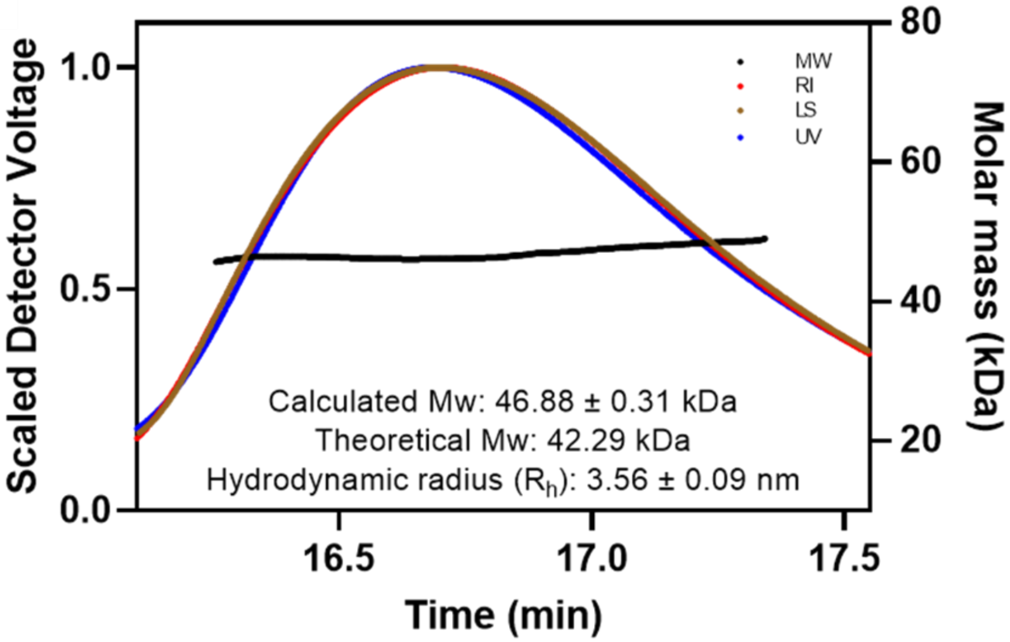


Figure 5.5. The MALS trace of the $NCL_{290-652}$ peak with the molecular weight (MW) (black), refractive index (RI) (red), light scattering (LS) (brown), and UV absorption (UV) (blue). These were collectively aligned after data collection and allowed for the determination of the MW, as well as the R_h.

5.4.3 Structural evaluation of the TRs and $NCL_{290-652}$ using SAXS

To further characterize the solution structure of the ZIKV TRs as well as $NCL_{290-652}$, we performed SAXS. SAXS in general allows biomolecular researchers to explore nanoscale structural characteristics of biological molecules in solution [77]. Due to the motion of particles tumbling in solution as they are passing through the X-ray source, the scattered radiation that can be collected and rendered accurately for structure determination between ~0.1° - 10° radially from the X-ray path, as this provides the most contrast against buffer scattering of the surrounding medium [78]. Prior to the sample passing through an X-ray beam, it is first subjected to SEC to separate any oligomers or degraded product, thus giving reliability in monodisperse data collection [78]. The samples were sent to the Diamond light source in Oxfordshire, England, and the raw data was sent back to our research group for analysis. Post data collection, the raw data was plotted using the relative scattering intensity as a function of scattering angle, which was sample is homogeneous, and not showing signs of aggregation for $NCL_{290-652}$ and the TRs (Fig. 5.7A $NCL_{290-652}$ data). Following this, a Guinier analysis was performed on $NCL_{290-652}$ and the TRs, yielding a radius of gyration (R_g) - the average root mean squared radius from the center of the mass for the respective biomolecules. This was determined to be 3.20 ± 0.01 nm for $NCL_{290-652}$, and the 5’ and 3’ TRs at 4.40 ± 0.09 nm, and 8.40 ± 0.21 nm, respectively (Fig. 5.7B $NCL_{290-652}$ data). Next, a dimensionless Kratky plot for determining the general shape of each species was performed and showing that all samples were folded (globular for $NCL_{290-652}$ and extended for the TRs) appropriately for the downstream development of low-resolution modelling (Fig. 5.7C). Finally, the data sets were evaluated using a paired-distance distribution analysis (P(r)) in which the reciprocal space data was converted into real-space electron density using indirect Fourier transformation [79], collectively yielding a D_{max} value, which is the maximum distance dimension of the species (Fig. 5.7D for $NCL_{290-652}$ data). These showed $NCL_{290-652}$ at 10.2 nm, ZIKV 5’ at 13.0 nm, and 3’ TR at 27 nm. Additionally, the P(r) also allows for the determination of real-space data (R_g) representing the actual distances within the molecule, which indicated good agreement with the Guinier (R_g) being at 3.2 ± 0.07 nm for $NCL_{290-652}$ With the combination of this data, and the ~50 low-resolution models generated for $NCL_{290-652}$ and the TRs, solution biophysical elucidation of the species is presented in Fig. 5.8 and gives clear indications of their 3-dimensional solution properties, laying the foundation for future targeting methods when factoring in size and shape of each species.

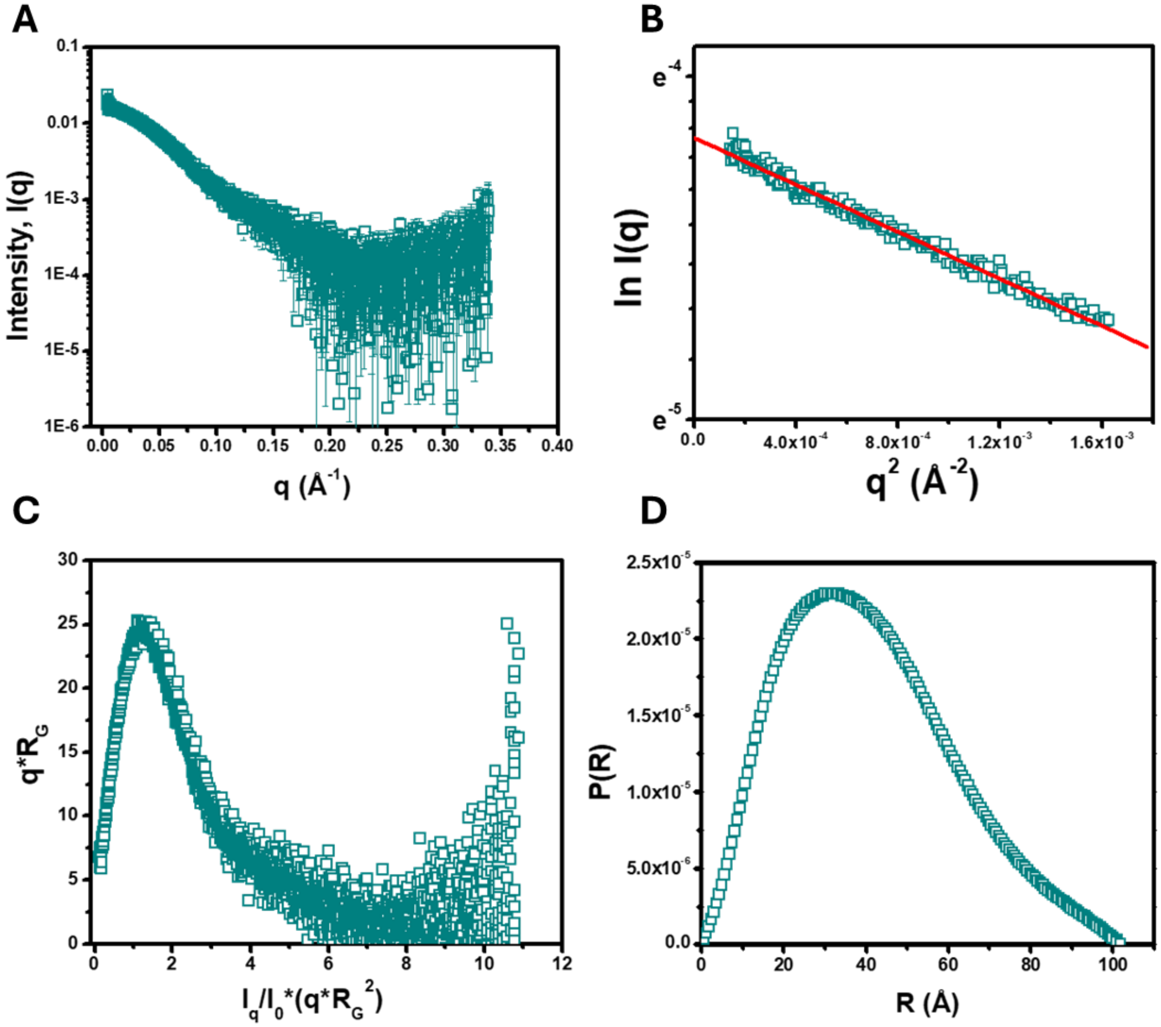


Figure 5.6. The SAXS data analysis plots for NCL$_{290-652}$. Panel **A** represents subtracted and merged data of the NCL$_{290-652}$ scattering representing the scattering intensity as a function of scattering angle (q = 4πsinΘ/λ). Panel **B** is the Guinier plot that determines the R_g as well as the purity of the overall sample. Panel **C** A Dimensionless Kratky plot of NCL$_{290-652}$ indicating globularity of the RRMs [80]. Panel **D** the determined paired-distance distribution plot that allowed the determination of the D_{max}, as well as the real-space R_g.

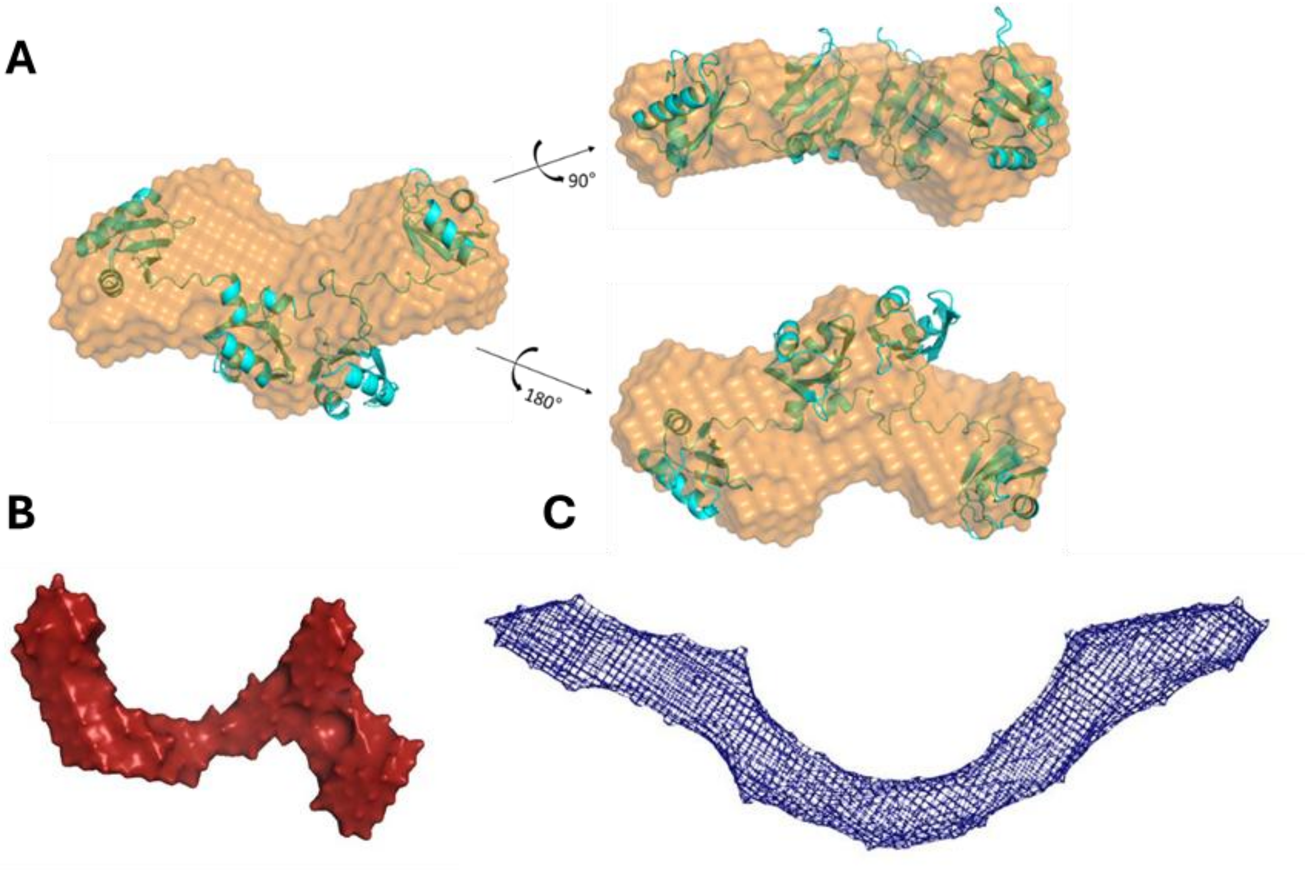


Figure 5.7. SAXS 3-dimensional models were produced for Panel **A** NCL$_{290-652}$, Panel **B** ZIKV 5' TR, and Panel **C**, the ZIKV 3' TR. The models were developed using *ab initio* model reconstruction. Furthermore, using AlphaFold of NCL$_{290-652}$, we fit the predicted high-resolution data (cyan) with the determined SAXS model and show a relative predicted representation of NCL$_{290-652}$ in solution.

5.4.4 The ZIKV TRs interact with high affinity to NCL$_{290-652}$

To retrieve and investigate kinetic elucidation if the ZIKV TRs and NCL$_{290-652}$ were directly interacting with each other based on the pulldown assay results, we performed electrophoretic mobility shift assays (EMSA) as well as microscale thermophoresis (MST). Furthermore, we aimed to understand if there was kinetic variation across the 3 general regions within the 3' TR, namely, the xrRNAs, the dumbbell region, and the 3' stem-loop. Collecting insights into this can provide further available options to directly target specific regions of the TRs that could be critical for

ZIKV to survive [81-83]. We first wanted to investigate if the TRs interacted with $NCL_{290-652}$ in general, therefore, 6% of native PAGE EMSAs were run with the 5' and 3' TR and $NCL_{290-652}$. EMSAs can provide a researcher with direct visual evidence if an interaction is occurring between two or more species [84]. However, extrapolating any dissociation constants is not valid, as the environment the species are run in is not completely physiologically relevant, as there are no salts present within the running buffer, and the species are in non-equilibrium conditions [84]. Nonetheless, EMSA does remain a strong tool for identifying that an interaction is occurring. It's also worth noting that the absence of binding on an EMSA does not conclude that the species do not interact, as many interactions require salts for the interaction to occur [84].

Based on the EMSA result (Fig. 5.8) we hypothesize that the ZIKV TRs and $NCL_{290-652}$ are interacting with each other, as there is a shift of nucleic acid species in the PAGE. This occurs due to a sudden change in mass, shape, and/or charge [84]. This shift is hypothesized to be $NCL_{290-652}$ interacting with the ZIKV TRs, as a "caging effect" is occurring on bands that are shifted higher, which is indicative of binding [85]. Furthermore, there are differences in shifting between the 5' and 3' TRs. On the 5' TR EMSA we see shifting starting to occur between 234 and 468 nM of $NCL_{290-652}$, and between 468 and 937 nM for the 3' TR. It is unclear as to why there is more band streaking occurring on the 3' vs. the 5' TR, but it could be hypothesized that it is binding to multiple sites on the 3' TR all with close but varying K_ds, and as well, weaker binding events at that specific concentration. Moreover, we investigated this further using MST and an RNA tiling approach, wherein we determined the kinetic differences between the various regions of the 3' TR.

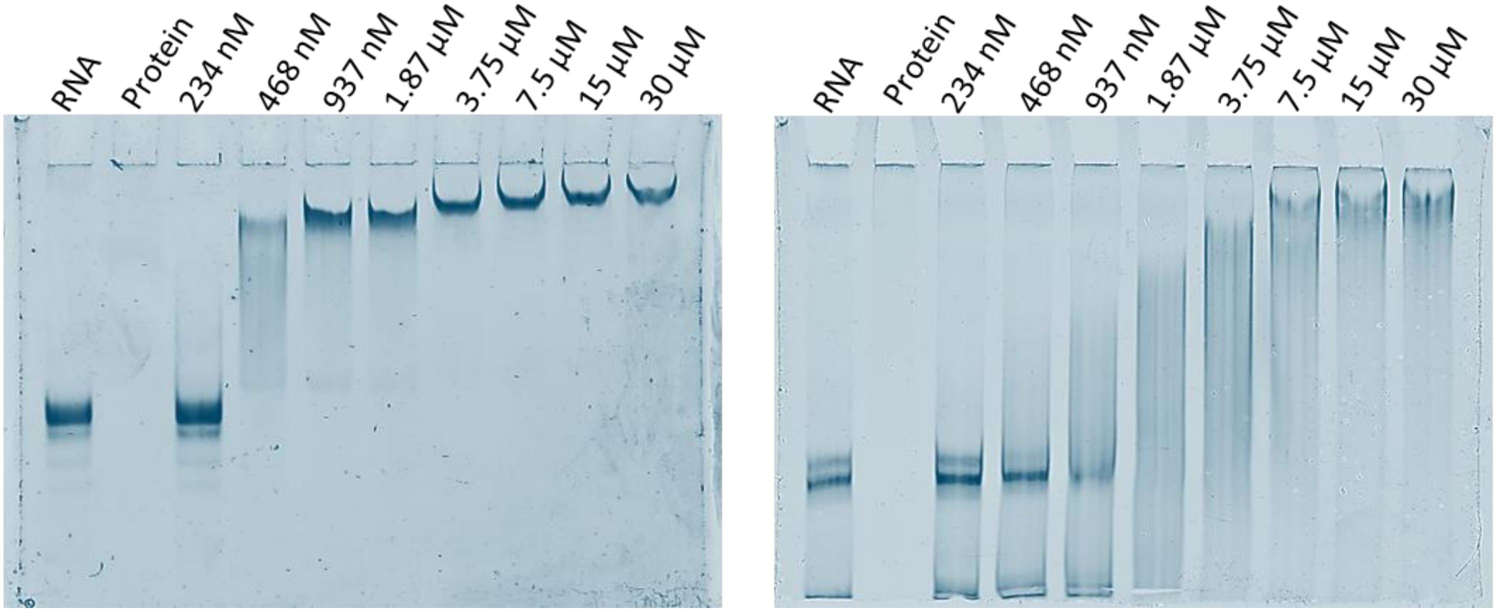
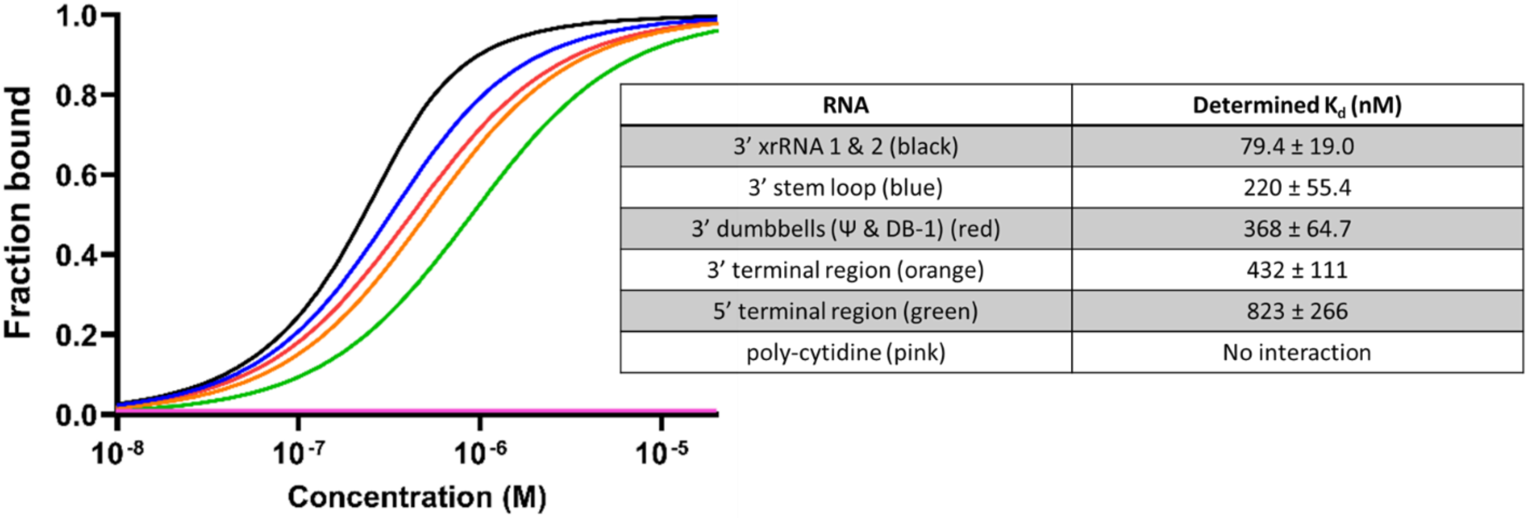


Figure 5.8. 6% native PAGEs of the 5’ (left) and 3’ (right) TRs with NCL$_{290\text{-}652}$. The 5’ TR shows a stark “caging effect” compared to the 3’ TR [85].

With the evidence of binding mentioned above with the EMSAs, we sought to perform MST in order to obtain a quantitative understanding of the kinetics of NCL$_{290\text{-}652}$ and the ZIKV TRs, as well as the specific sections of the 3’ TR using the RNA tiling approach. This powerful technique allows researchers to retrieve (not limited to) a dissociation constant (K_d) of two species, including nucleic acids, proteins, carbohydrates, lipids, and small molecules, all together in a wide range of K_d’s (pM - µM) and complex buffer mixtures including cell lysate [86, 87]. We performed MST in simplified conditions – with pure, fluorescently labeled (Alexa® 488) RNA, and NCL$_{290\text{-}652}$. A 2-fold serial dilution of NCL$_{290\text{-}652}$ was performed, whilst keeping the labelled RNA concentration constant (n=3) (Fig. 5.9). An excitation laser was applied to the samples, causing fluorescence of the RNA, and an infrared laser was shone onto the sample, causing a thermal gradient [86, 87]. The thermal diffusion of the fluorescence is tracked over time and analyzed by plotting the change in fluorescence as a function of time [86, 87]. Data collection was performed across the ZIKV 5’ and 3’ TRs, as well as the 3’ xrRNA 1 & 2, 3’ dumbbells, and the 3’ stem-loop. Previous literature indicated that NCL does not interact with large stretches of cytidine, therefore, we employed labelled poly-cytidine as a negative control [88]. We also compared the dissociation constants of the RRM and full-length NCL in previous literature, which altogether shows K_d’s between 75 – 700 nM [89, 90]. The ZIKV 5’ TR was run, and a K_d of 823 nM was retrieved, which was the highest K_d out of all the RNAs that were measured. Following that, the 3’ TR was measured with $NCL_{290-652}$ and a K_d of 432 nM was determined. Moreover, we then investigated the 3 sections of the 3’ TR - xrRNAs, the dumbbell region, and the 3’ stem-loop in order to gain insight if there were any potential preferences for one section over another. Interestingly, the 3’ dumbbells had a K_d of 368 nM, the 3’ stem-loop at 220 nM, and the xrRNA 1 & 2 at 79.4 nM (Fig. 5.9).



RNA	Determined K_d (nM)
3’ xrRNA 1 & 2 (black)	79.4 ± 19.0
3’ stem loop (blue)	220 ± 55.4
3’ dumbbells (Ψ & DB-1) (red)	368 ± 64.7
3’ terminal region (orange)	432 ± 111
5’ terminal region (green)	823 ± 266
poly-cytidine (pink)	No interaction

Figure 5.9. A summary of the MST data collection of the 5’ and 3’ TRs, as well as the 3 sections of the 3’ TR (xrRNA 1 & 2, dumbbells, and 3’ stem-loop) (n=3). Each colour on the MST plots corresponds to the colour labels on the table of each transcript. Pink indicates the poly-cytidine runs that showed no interaction occurring, indicating binding preference and specificity to specific sequences/structures of RNA.

From these results, it is still unclear why there is a kinetic preference towards the xrRNAs 1 & 2, but more research is required to elucidate these results. One could hypothesize that in general, a smaller RNA transcript (i.e. the tiled RNA) could provide more steric freedom, allowing $NCL_{290-652}$ more access to these individual sections, but it remains clear that from a physiological perspective, this section is part of an entire viral RNA genome, therefore, the kinetics may differ [91-93]. It could be hypothesized that the preferential kinetics of $NCL_{290-652}$ to the xrRNA 1 & 2, could be a critical insight into the virus's need to produce sub-genomic flaviviral RNA (sfRNA) for efficient viral replication [94-97]. The xrRNAs are involved in RNase resistance during viral amplification in the cell, consequently, sfRNA in the cell (3' TR strand accumulation) [94-97]. These sfRNAs allow for the entire kinetic landscape of the cell to change by the sfRNA "protein sponging" host proteins, favouring viral replication more [94-97]. Moreover, we could hypothesize that NCLs involvement in sfRNA production is potentially providing aid to viral replication in some way. More investigation is needed as to why this preferential binding is occurring. Overall, the data provided context that NCL bound directly to the ZIKV TR RNA in the pulldown assay, though more physiological relevance is needed, and we turned to cell biology to understand if NCL is critical for ZIKV replication.

5.4.5 siRNA knockdown of NCL in Huh-7.5 cells prevents ZIKV virion formation

In collaboration with the Chatel-Chaix research group at the University of Laval, we performed siRNA knockdowns of NCL in the Huh-7.5 hepatocarcinoma human cell line. Looking to viral phylogeny, the *Zika virus*, as well as the *Dengue virus* share a close common ancestor in within the flaviviral family compared to other flaviviral species [98-100]. As previously mentioned, these viruses cause ADE, are in the same geographical locations, and can be coinfected in an individual mosquito [101-103]. Because of this genetic similarity, we aim to investigate if NCL is also critical for DENV amplification by performing NCL knockdowns in tandem with ZIKV. This was altogether achieved by transfection of siRNA that targets NCL mRNA transcripts, thus preventing NCL translation. To ensure that NCL was knocked down in the Huh-7.5 cells for both ZIKV and DENV, western blots were performed to show non-detectable amounts of NCL compared to controls, such as the viral NS3 protein, human actin, and the typical expression pattern of NCL itself (Fig. 5.10). The western blotting results show a stark contrast between the NCL wild-type expression and the transfected siRNA cells. Though a non-detectable amount of NCL is occurring, the NS3 viral protein for ZIKV and DENV have remained relatively the same, which is worth noting, as Figure 7.11 shows noteworthy results from the NS3 intensity on the blot, and the lack of viral transcript detected.

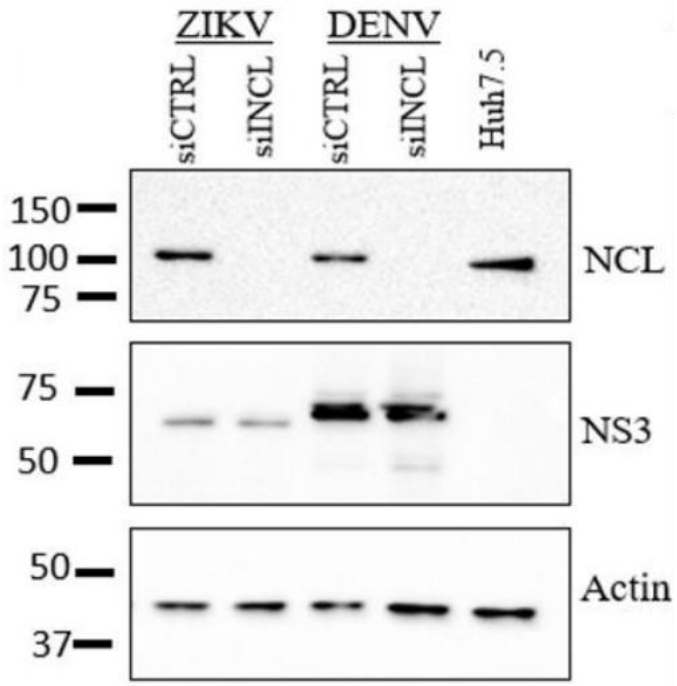


Figure 5.10. Western blot results from the siRNA knockdowns of NCL in DENV and ZIKV using Huh-7.5 cells, together indicating non-detectable amounts of NCL in the cell. Actin was used as a positive control, as well as the viral NS3 protein, indicating that the virus is replicating inside cells.

A plaque assay was performed which used 10-fold serial dilutions of a stock virus sample titrated onto a cellular monolayer in order to quantitatively measure the amount of virions infecting the cells based on the number of plaques that form [104]. The plaques are counted and reported in plaque-forming units (PFU) per millilitre of cells [104]. Figure 7.12 shows the change in PFU/mL, otherwise known as the viral titer, as a function of NCL siRNA untreated vs. treated cells. Moreover, from this assay, a drastic change in the PFUs in both ZIKV and DENV, with ZIKV having a stark change in the number of functional virions forming. However, we have not elucidated as to why we see this difference in PFU amount between ZIKV and DENV, but one could hypothesize that NCL is more critical for the formation of functional virions in ZIKV vs. DENV. Furthermore, a t-test was performed to retrieve statistical significance on the plaque assay, indicating two asterisks, meaning that the p-value of this test is <0.01, which is statistically significant [105]. It is important to note that a decrease in the number of plaques indicates a decrease in the number of infectious virus particles formed, but does not indicate total virus particles, as there could be non-infectious particle formation, therefore, conclusively of lethality to the virus is still under debate [106, 107]. Nonetheless, these results form a foundation for an alternative way to potentially target ZIKV and/or DENV. Moreover, we also measured the amount of viral RNA present relative to the amount of PFUs and elucidated a statistically significant decrease in both ZIKV and DENV RNA (<0.05 and <0.01, respectively). We hypothesize that this result could be correlated with the formation of functional virions as well, but, with the NS3 western blot result, there is no significant decrease in the amount of NS3 that is translated.

Because NS3 is altogether part of the polyprotein that is collectively translated, it could be hypothesized that since the NS3 is at the same intensity, all other viral proteins are at the same intensity, but more research is required. If the viral transcripts are decreased during NCL knockdown, then it can be assumed that viral translation is not impacted by this, therefore, there is an impact occurring somewhere else in the viral lifecycle. For example, potentially the TRs are involved with NCL, and cyclization could be stifled, allowing translation to occur relatively optimally, but decreasing cyclization base pairing, which causes a decrease in viral transcripts generated by NS5 [19]. Furthermore, a variant hypothesis could be that NCL aids in viral packaging, and/or transcription in general, as with the variation in the number of transcripts, one could hypothesize potential abortive transcripts, leading to non-functional virus particles due to NCL not being present to aid in those processes. It could be hypothesized as well that NCL is disrupting an alternative pathway in the cell that is having a downstream effect on virion formation and/or viral transcription. In conclusion, more elucidation is required to understand the mechanisms of this constraining effect of PFUs and viral transcripts in the absence of NCL.

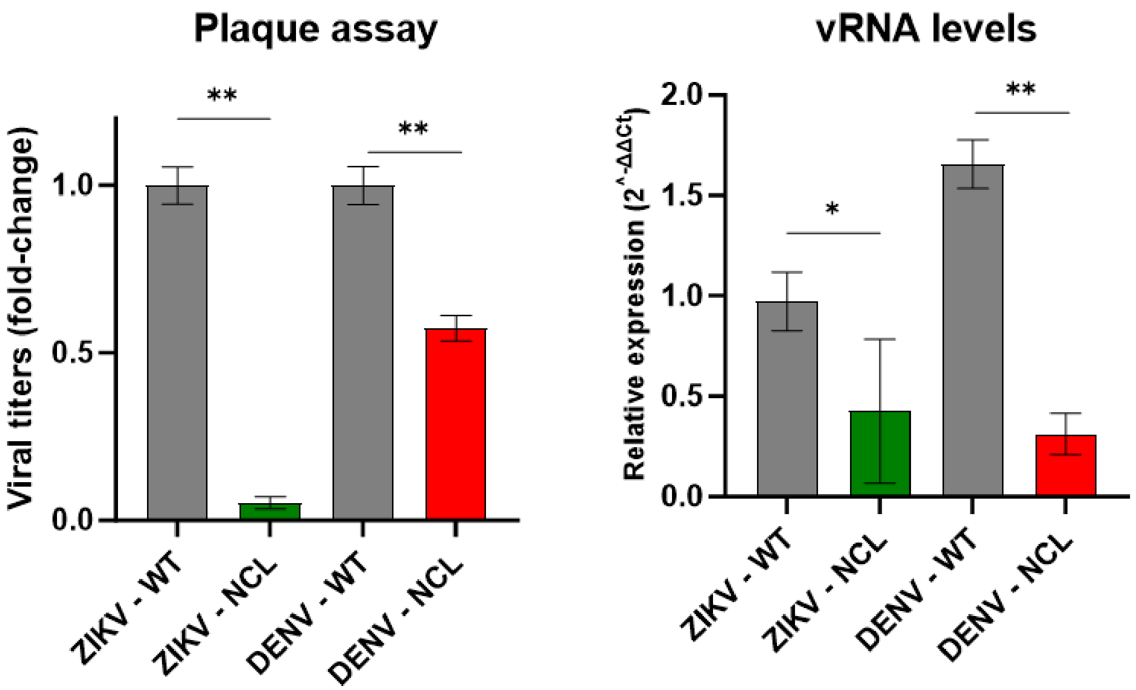


Figure 5.11. A plaque assay and viral qPCR quantification to understand if NCL knockdowns have an impact on the formation of virions, and/or viral RNA levels (n=3). Grey plots indicate the wild-type controls of Huh-7.5 without NCL knockdown, showing a relative standard of virion formation. Green is ZIKV assays, and red is DENV assays.

5.5 Conclusion

ZIKV and DENV continue to be a significant issue globally, as they collectively cause ~400 million infections worldwide, and due to ADE, the vaccination options are limited. No antivirals have been developed for either of these viruses and with ADE occurring with the subsequent challenge of the virus on an individual, causing hemorrhagic fever and death, urgency is required to combat these viruses and lower the public health impact these infectious agents cause. We aimed to provide a means to target both viruses by evaluating the regions of these viruses that are conserved and critical for viral replication. The highest conserved area within the ZIKV and DENV genomes are the TRs themselves, because of their previously mentioned criticality. There are no known antivirals developed for any virus that targets the viral non-coding RNA-human RNA binding protein interface, therefore, this elucidation is a novel approach to antiviral discovery. We sought to identify the human proteins that interact with these conserved regions via pulldown assay, and NCL was a hypothesized potential direct binding partner. We identified that NCL is a direct RNA-binding protein to the TRs and performed structural characterization on the TRs as well as NCL. Due to potential oligomerization and/or liquid-liquid phase separation potential, characterizing these interfaces through biophysical means proves to be difficult, and more optimization is required.

We kinetically characterized these interactions via EMSA and MST and retrieved nanomolar affinities that coincide with previous literature on NCLs' affinity to RNA. Moreover, we showed that there is preferential binding with the xrRNA 1 & 2 of ZIKV, which serve for the generation of sfRNAs. These results collectively show that understanding specific pathways and mechanisms of NCL during ZIKV and DENV replication is needed, more importantly regarding sfRNA production because of the MST results. The cell biology assays revealed that the PFUs and viral RNA are significantly stifled when NCL is knocked down in Huh-7.5 cells, giving promise that NCL could play a critical role in the transcription of the virus, and/or the formation of functional virions. Altogether, this evidence provides a means to further investigate NCLs' role in ZIKV and DENV replication, as well as explore available drugs or small molecules that impact NCL enough to disrupt the binding interfaces with the ZIKV and DENV RNA, providing a potential means of treatment options. Additionally, if this avenue of treatment provides an option for health providers, then further investigation into multi-viral treatments could be an option, as the TRs across flaviviruses possess many identical structures/sequences that NCL could interact with [108-111]. This would effectively allow for multiple flaviviruses to be targeted that all impact agriculture and/or public health, thus providing a novel channel to empower healthcare workers and agricultural professionals by lowering the impact of these infectious agents.

5.6 References

1. Sinden, R.E., *Malaria, mosquitoes and the legacy of Ronald Ross.* Bull World Health Organ, 2007. **85**(11): p. 894-6.
2. Xie, S., et al., *AXL, an Important Host Factor for DENV and ZIKV Replication.* Front Cell Infect Microbiol, 2021. **11**: p. 575346.
3. Ramirez Alvarez, C., et al., *The endogenous cellular protease inhibitor SPINT2 controls SARS-CoV-2 viral infection and is associated to disease severity.* PLOS Pathogens, 2021. **17**(6): p. e1009687.
4. Azevedo, R.S.S., et al., *In situ immune response and mechanisms of cell damage in central nervous system of fatal cases microcephaly by Zika virus.* Sci Rep, 2018. **8**(1): p. 1.
5. Leone, T., et al., *Web-based searching for abortion information during health emergencies: a case study of Brazil during the 2015/2016 Zika outbreak.* Sex Reprod Health Matters, 2021. **29**(1): p. 1883804.
6. Hindle, S., et al., *Zika virus infection during pregnancy and vertical transmission: case reports and peptide-specific cell-mediated immune responses.* Arch Virol, 2024. **169**(2): p. 32.
7. Koenig, M.R., et al., *Vertical transmission of African-lineage Zika virus through the fetal membranes in a rhesus macaque (Macaca mulatta) model.* PLOS Pathogens, 2023. **19**(8): p. e1011274.
8. Izmirly, A.M., et al., *Challenges in Dengue Vaccines Development: Pre-existing Infections and Cross-Reactivity.* Front Immunol, 2020. **11**: p. 1055.
9. Hou, J., W. Ye, and J. Chen, *Current Development and Challenges of Tetravalent Live-Attenuated Dengue Vaccines.* Frontiers in Immunology, 2022. **13**.
10. Pintado Silva, J. and A. Fernandez-Sesma, *Challenges on the development of a dengue vaccine: a comprehensive review of the state of the art.* Journal of General Virology, 2023. **104**(3).
11. Xu, L., et al., *Antibody dependent enhancement: Unavoidable problems in vaccine development.* Adv Immunol, 2021. **151**: p. 99-133.
12. Al Fayez, N., et al., *Recent Advancement in mRNA Vaccine Development and Applications.* Pharmaceutics, 2023. **15**(7).
13. Baz, M. and G. Boivin, *Antiviral Agents in Development for Zika Virus Infections.* Pharmaceuticals (Basel), 2019. **12**(3).
14. Zhang, N., et al., *Identification of novel anti-ZIKV drugs from viral-infection temporal gene expression profiles.* Emerg Microbes Infect, 2023. **12**(1): p. 2174777.
15. Jackman, J.A., et al., *Therapeutic treatment of Zika virus infection using a brain-penetrating antiviral peptide.* Nature Materials, 2018. **17**(11): p. 971-977.
16. Göertz, G.P., et al., *Functional RNA during Zika virus infection.* Virus Res, 2018. **254**: p. 41-53.
17. Sanford, T.J., et al., *Circularization of flavivirus genomic RNA inhibits de novo translation initiation.* Nucleic Acids Res, 2019. **47**(18): p. 9789-9802.
18. Liu, Z.-Y., et al., *Viral RNA switch mediates the dynamic control of flavivirus replicase recruitment by genome cyclization.* eLife, 2016. **5**: p. e17636.

19. Khromykh, A.A., et al., *Essential role of cyclization sequences in flavivirus RNA replication.* J Virol, 2001. **75**(14): p. 6719-28.
20. Mrozowich, T., et al., *Investigating RNA-RNA interactions through computational and biophysical analysis.* Nucleic Acids Res, 2023. **51**(9): p. 4588-4601.
21. Ochsenreiter, R., I.L. Hofacker, and M.T. Wolfinger, *Functional RNA Structures in the 3'UTR of Tick-Borne, Insect-Specific and No-Known-Vector Flaviviruses.* Viruses, 2019. **11**(3): p. 298.
22. de Chassey, B., et al., *Virus-host interactomics: new insights and opportunities for antiviral drug discovery.* Genome Medicine, 2014. **6**(11): p. 115.
23. Idrees, S., et al. *Exploring Viral–Host Protein Interactions as Antiviral Therapies: A Computational Perspective*. Microorganisms, 2024. **12**, DOI: 10.3390/microorganisms12030630.
24. Liu, X., et al., *SARS-CoV-2-host proteome interactions for antiviral drug discovery.* Mol Syst Biol, 2021. **17**(11): p. e10396.
25. Ma-Lauer, Y., et al., *Virus-host interactomes--antiviral drug discovery.* Curr Opin Virol, 2012. **2**(5): p. 614-21.
26. Zhang, S., et al., *Comparison of viral RNA–host protein interactomes across pathogenic RNA viruses informs rapid antiviral drug discovery for SARS-CoV-2.* Cell Research, 2022. **32**(1): p. 9-23.
27. Noorbakhsh, F., et al., *Zika Virus Infection, Basic and Clinical Aspects: A Review Article.* Iran J Public Health, 2019. **48**(1): p. 20-31.
28. Emmott, E. and I. Goodfellow, *Identification of protein interaction partners in mammalian cells using SILAC-immunoprecipitation quantitative proteomics.* J Vis Exp, 2014(89).
29. Tajrishi, M.M., R. Tuteja, and N. Tuteja, *Nucleolin.* Communicative & Integrative Biology, 2011. **4**(3): p. 267-275.
30. Tuteja, R. and N. Tuteja, *Nucleolin: a multifunctional major nucleolar phosphoprotein.* Crit Rev Biochem Mol Biol, 1998. **33**(6): p. 407-36.
31. Kumar, D., S. Broor, and M.S. Rajala, *Interaction of Host Nucleolin with Influenza A Virus Nucleoprotein in the Early Phase of Infection Limits the Late Viral Gene Expression.* PLoS One, 2016. **11**(10): p. e0164146.
32. Terrier, O., et al., *Nucleolin interacts with influenza A nucleoprotein and contributes to viral ribonucleoprotein complexes nuclear trafficking and efficient influenza viral replication.* Scientific Reports, 2016. **6**(1): p. 29006.
33. Chaudhry, U.A., et al. *Nucleolin: Role in Bacterial and Viral Infections*. 2018.
34. Tonello, F., M.L. Massimino, and C. Peggion, *Nucleolin: a cell portal for viruses, bacteria, and toxins.* Cellular and Molecular Life Sciences, 2022. **79**(5): p. 271.
35. Luige, J., et al., *Predicting nuclear G-quadruplex RNA-binding proteins with roles in transcription and phase separation.* Nature Communications, 2024. **15**(1): p. 2585.
36. Dash, S., et al., *rRNA transcription is integral to phase separation and maintenance of nucleolar structure.* PLoS Genet, 2023. **19**(8): p. e1010854.
37. Guillen-Chable, F., et al., *Phase Separation of Intrinsically Disordered Nucleolar Proteins Relate to Localization and Function.* Int J Mol Sci, 2021. **22**(23).
38. Gemmill, D.L., et al., *The 3' terminal region of Zika virus RNA contains a conserved G-quadruplex and is unfolded by human DDX17.* Biochem Cell Biol, 2024. **102**(1): p. 96-105.

39. Lotthammer, J.M., et al., *Direct prediction of intrinsically disordered protein conformational properties from sequence.* Nature Methods, 2024. **21**(3): p. 465-476.
40. Lapeyre, B., H. Bourbon, and F. Amalric, *Nucleolin, the major nucleolar protein of growing eukaryotic cells: an unusual protein structure revealed by the nucleotide sequence.* Proc Natl Acad Sci U S A, 1987. **84**(6): p. 1472-6.
41. Serin, G., et al., *Two RNA-binding Domains Determine the RNA-binding Specificity of Nucleolin *.* Journal of Biological Chemistry, 1997. **272**(20): p. 13109-13116.
42. Perez-Iratxeta, C. and M.A. Andrade-Navarro, *K2D2: Estimation of protein secondary structure from circular dichroism spectra.* BMC Structural Biology, 2008. **8**(1): p. 25.
43. Rosignoli, S. and A. Paiardini, *Boosting the Full Potential of PyMOL with Structural Biology Plugins.* Biomolecules, 2022. **12**(12).
44. Arumugam, S., et al., *Solution structure of the RBD1,2 domains from human nucleolin.* Journal of biomolecular NMR, 2010. **47**(1): p. 79-83.
45. Thakur, M., et al., *EMBL's European Bioinformatics Institute (EMBL-EBI) in 2022.* Nucleic Acids Research, 2023. **51**(D1): p. D9-D17.
46. Europe, P.D.B.i. *PDB 2fc8 structure summary: Solution structure of the RRM_1 domain of NCL protein*. 2006; Available from: https://www.ebi.ac.uk/pdbe/entry/pdb/2fc8.
47. Europe, P.D.B.i. *PDB 2fc9 structure summary: Solution structure of the RRM_1 domain of NCL protein*. 2006; Available from: https://www.ebi.ac.uk/pdbe/entry/pdb/2fc9.
48. Amartely, H., et al., *Coupling Multi Angle Light Scattering to Ion Exchange chromatography (IEX-MALS) for protein characterization.* Scientific Reports, 2018. **8**(1): p. 6907.
49. Zhao, H., P.H. Brown, and P. Schuck, *On the distribution of protein refractive index increments.* Biophys J, 2011. **100**(9): p. 2309-17.
50. Meier, M., et al., *Structure and hydrodynamics of a DNA G-quadruplex with a cytosine bulge.* Nucleic Acids Res, 2018. **46**(10): p. 5319-5331.
51. Franke, D., et al., *ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions.* J Appl Crystallogr, 2017. **50**(Pt 4): p. 1212-1225.
52. Panjkovich, A. and D.I. Svergun, *CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data.* Bioinformatics, 2018. **34**(11): p. 1944-1946.
53. Manalastas-Cantos, K., et al., *ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis.* J Appl Crystallogr, 2021. **54**(Pt 1): p. 343-355.
54. Putnam, C.D., *Guinier peak analysis for visual and automated inspection of small-angle X-ray scattering data.* J Appl Crystallogr, 2016. **49**(Pt 5): p. 1412-1419.
55. Burke, J.E. and S.E. Butcher, *Nucleic acid structure characterization by small angle X-ray scattering (SAXS).* Curr Protoc Nucleic Acid Chem, 2012. **Chapter 7**: p. Unit7.18.
56. Rambo, R.P. and J.A. Tainer, *Improving small-angle X-ray scattering data for structural analyses of the RNA world.* Rna, 2010. **16**(3): p. 638-46.
57. Fang, X., et al., *Small-angle X-ray scattering: a bridge between RNA secondary structures and three-dimensional topological structures.* Curr Opin Struct Biol, 2015. **30**: p. 147-160.

58. Rambo, R.P. and J.A. Tainer, *Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law.* Biopolymers, 2011. **95**(8): p. 559-71.
59. Semenyuk, A. and D. Svergun, *GNOM–a program package for small-angle scattering data processing.* Journal of applied crystallography, 1991. **24**(5): p. 537-540.
60. Svergun, D., *Determination of the regularization parameter in indirect-transform methods using perceptual criteria.* Journal of applied crystallography, 1992. **25**(4): p. 495-503.
61. Svergun, D.I., *Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing.* Biophys J, 1999. **76**(6): p. 2879-86.
62. Franke, D. and D.I. Svergun, *DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering.* J Appl Crystallogr, 2009. **42**(Pt 2): p. 342-346.
63. López-Méndez, B., et al., *Reproducibility and accuracy of microscale thermophoresis in the NanoTemper Monolith: a multi laboratory benchmark study.* Eur Biophys J, 2021. **50**(3-4): p. 411-427.
64. Kitagawa, S., et al., *Elucidation of the role of nucleolin as a cell surface receptor for nucleic acid-based adjuvants.* npj Vaccines, 2022. **7**(1): p. 115.
65. Fujiki, H., T. Watanabe, and M. Suganuma, *Cell-surface nucleolin acts as a central mediator for carcinogenic, anti-carcinogenic, and disease-related ligands.* J Cancer Res Clin Oncol, 2014. **140**(5): p. 689-99.
66. Mazeaud, C., et al., *Zika virus remodels and hijacks IGF2BP2 ribonucleoprotein complex to promote viral replication organelle biogenesis*. 2023, Cold Spring Harbor Laboratory.
67. Mrozowich, T., et al., *Nanoscale Structure Determination of Murray Valley Encephalitis and Powassan Virus Non-Coding RNAs.* Viruses, 2020. **12**(2).
68. Nelson, C., et al., *Human DDX3X Unwinds Japanese Encephalitis and Zika Viral 5' Terminal Regions.* Int J Mol Sci, 2021. **22**(1).
69. Nelson, C.R., et al., *Human DDX17 Unwinds Rift Valley Fever Virus Non-Coding RNAs.* Int J Mol Sci, 2020. **22**(1).
70. Baldwin, R.L. and G.D. Rose, *Is protein folding hierarchic? I. Local structure and peptide folding.* Trends in Biochemical Sciences, 1999. **24**(1): p. 26-33.
71. Chen, Y., et al., *Protein folding: then and now.* Arch Biochem Biophys, 2008. **469**(1): p. 4-19.
72. Childers, M.C., C.-L. Towse, and V. Daggett, *The effect of chirality and steric hindrance on intrinsic backbone conformational propensities: tools for protein design.* Protein Engineering, Design and Selection, 2016. **29**(7): p. 271-280.
73. Velours, C., et al., *Determination of the Absolute Molar Mass of [Fe-S]-Containing Proteins Using Size Exclusion Chromatography-Multi-Angle Light Scattering (SEC-MALS).* Biomolecules, 2022. **12**(2).
74. Clobes, M.L., E.I. Kozliak, and A. Kubátová, *Advancing Molecular Weight Determination of Lignin by Multi-Angle Light Scattering.* Polymers, 2024. **16**(4): p. 477.
75. Andrianova, A.A., et al., *Size exclusion chromatography of lignin: The mechanistic aspects and elimination of undesired secondary interactions.* Journal of Chromatography A, 2018. **1534**: p. 101-110.
76. Stetefeld, J., S.A. McKenna, and T.R. Patel, *Dynamic light scattering: a practical guide and applications in biomedical sciences.* Biophys Rev, 2016. **8**(4): p. 409-427.

77. Di Cola, E., I. Grillo, and S. Ristori, *Small Angle X-ray and Neutron Scattering: Powerful Tools for Studying the Structure of Drug-Loaded Liposomes.* Pharmaceutics, 2016. **8**(2).
78. Malaby, A.W., et al., *Methods for analysis of size-exclusion chromatography-small-angle X-ray scattering and reconstruction of protein scattering.* J Appl Crystallogr, 2015. **48**(Pt 4): p. 1102-1113.
79. Kikhney, A.G. and D.I. Svergun, *A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins.* FEBS Lett, 2015. **589**(19 Pt A): p. 2570-7.
80. Bernadó, P. and D.I. Svergun, *Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering.* Molecular BioSystems, 2012. **8**(1): p. 151-167.
81. Claas, A.M., et al., *Viral Kinetics Model of SARS-CoV-2 Infection Informs Drug Discovery, Clinical Dose, and Regimen Selection.* Clinical Pharmacology & Therapeutics, 2024. **n/a**(n/a).
82. Ahmed, S., et al., *Investigating the binding affinity, interaction, and structure-activity-relationship of 76 prescription antiviral drugs targeting RdRp and Mpro of SARS-CoV-2.* J Biomol Struct Dyn, 2021. **39**(16): p. 6290-6305.
83. Singh, S., et al., *Identification of multi-targeting natural antiviral peptides to impede SARS-CoV-2 infection.* Struct Chem, 2022: p. 1-16.
84. Hellman, L.M. and M.G. Fried, *Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions.* Nat Protoc, 2007. **2**(8): p. 1849-61.
85. Heffler, M.A., R.D. Walters, and J.F. Kugel, *Using electrophoretic mobility shift assays to measure equilibrium dissociation constants: GAL4-p53 binding DNA as a model system.* Biochem Mol Biol Educ, 2012. **40**(6): p. 383-7.
86. Magnez, R., C. Bailly, and X. Thuru, *Microscale Thermophoresis as a Tool to Study Protein Interactions and Their Implication in Human Diseases.* Int J Mol Sci, 2022. **23**(14).
87. Shao, W., et al., *Microscale thermophoresis as a powerful tool for screening glycosyltransferases involved in cell wall biosynthesis.* Plant Methods, 2020. **16**(1): p. 99.
88. González, V., et al., *Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein.* J Biol Chem, 2009. **284**(35): p. 23622-35.
89. Masuzawa, T. and T. Oyoshi, *Roles of the RGG Domain and RNA Recognition Motif of Nucleolin in G-Quadruplex Stabilization.* ACS Omega, 2020. **5**(10): p. 5202-5208.
90. Li, F., et al., *A water-soluble nucleolin aptamer-paclitaxel conjugate for tumor-specific targeting in ovarian cancer.* Nat Commun, 2017. **8**(1): p. 1390.
91. Sharma, D., et al., *The kinetic landscape of an RNA-binding protein in cells.* Nature, 2021. **591**(7848): p. 152-156.
92. Sinan, S., N.M. Appleby, and R. Russell, *Kinetic dissection of pre-crRNA binding and processing by CRISPR-Cas12a.* bioRxiv, 2023.
93. Nagasawa, R., et al., *Large-scale analysis of small molecule-RNA interactions using multiplexed RNA structure libraries.* Communications Chemistry, 2024. **7**(1): p. 98.
94. Funk, A., et al., *RNA Structures Required for Production of Subgenomic Flavivirus RNA.* Journal of Virology, 2010. **84**(21): p. 11407-11417.
95. Watkins, J.M. and J.M. Burke, *RNase L-induced bodies sequester subgenomic flavivirus RNAs and re-establish host RNA decay.* bioRxiv, 2024.
96. Liu, Y., W. Guan, and H. Liu, *Subgenomic Flaviviral RNAs of Dengue Viruses.* Viruses, 2023. **15**(12): p. 2306.

97. Slonchak, A., et al., *Structural analysis of 3’UTRs in insect flaviviruses reveals novel determinants of sfRNA biogenesis and provides new insights into flavivirus evolution.* Nature Communications, 2022. **13**(1): p. 1279.
98. Boldescu, V., et al., *Broad-spectrum agents for flaviviral infections: dengue, Zika and beyond.* Nature Reviews Drug Discovery, 2017. **16**(8): p. 565-586.
99. Wang, A., et al., *Zika virus genome biology and molecular pathogenesis.* Emerg Microbes Infect, 2017. **6**(3): p. e13.
100. Song, Y., et al., *Dengue and Zika Virus 5’ Untranslated Regions Harbor Internal Ribosomal Entry Site Functions.* mBio, 2019. **10**(2): p. 10.1128/mbio.00459-19.
101. Lin, D.C.-D., et al., *Co-infection of dengue and Zika viruses mutually enhances viral replication in the mosquito Aedes aegypti.* Parasites & Vectors, 2023. **16**(1): p. 160.
102. Azeredo, E.L., et al., *Clinical and Laboratory Profile of Zika and Dengue Infected Patients: Lessons Learned From the Co-circulation of Dengue, Zika and Chikungunya in Brazil.* PLoS Curr, 2018. **10**.
103. Chaves, B.A., et al., *Coinfection with Zika Virus (ZIKV) and Dengue Virus Results in Preferential ZIKV Transmission by Vector Bite to Vertebrate Host.* J Infect Dis, 2018. **218**(4): p. 563-571.
104. Baer, A. and K. Kehn-Hall, *Viral concentration determination through plaque assays: using traditional and novel overlay systems.* J Vis Exp, 2014(93): p. e52065.
105. *GraphPad Prism 10 User Guide - How to cite GraphPad Prism*. 0% Journal Article.
106. Jones, J.E., V. Le Sage, and S.S. Lakdawala, *Viral and host heterogeneity and their effects on the viral life cycle.* Nature Reviews Microbiology, 2021. **19**(4): p. 272-282.
107. Klasse, P.J., *Molecular determinants of the ratio of inert to infectious virus particles.* Prog Mol Biol Transl Sci, 2015. **129**: p. 285-326.
108. Chiu, W.-W., M. Kinney Richard, and W. Dreher Theo, *Control of Translation by the 5′- and 3′-Terminal Regions of the Dengue Virus Genome.* Journal of Virology, 2005. **79**(13): p. 8303-8315.
109. Markoff, L., *5'- and 3'-noncoding regions in flavivirus RNA.* Adv Virus Res, 2003. **59**: p. 177-228.
110. Ng, W.C., et al. *The 5′ and 3′ Untranslated Regions of the Flaviviral Genome*. Viruses, 2017. **9**, DOI: 10.3390/v9060137.
111. Fernández-Sanlés, A., et al., *Functional Information Stored in the Conserved Structural RNA Domains of Flavivirus Genomes.* Front Microbiol, 2017. **8**: p. 546.

Preface: ***<u>This chapter was published in the Journal of Biochemistry and Cell Biology in Canadian Science Publishing, Volume 102, Number 1, February 2024, and is in its standard journalistic format.</u>*** **Danielle Gemmill wrote the entire manuscript, and made all figures, and performed all experiments and data analysis except for the locARNA alignment and figure, and the MST DDX17/DDX3X interaction with the 3'TR. These experiments and figures were performed and made by Michael T. Wolfinger and Corey R. Nelson, respectively. Maulik D. Badmalia, Higor S. Pereira assisted in editing the manuscript, and providing insight into some of the experimentation. Liam Kerr assisted with the fluorimetry experiments.**

Chapter 6. The 3' terminal region of Zika virus RNA contains a conserved G-quadruplex and is unfolded by human DDX17

Danielle Gemmill[a#], Corey R. Nelson[a#], Maulik D. Badmalia[a], Higor S. Pereira[a], Liam Kerr[a], Michael T. Wolfinger[b,c], and Trushar R. Patel*[a,d,e]

[a]Alberta RNA Research and Training Institute & Department of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, AB T1K 3M4, Canada.

[b]Bioinformatics and Computational Biology, Faculty of Computer Science, University of Vienna, Währinger Strasse 29, 1090, Vienna, Austria

[c]Department of Theoretical Chemistry, University of Vienna, Währinger Strasse 17, 1090, Vienna, Austria

[d]Department of Microbiology, Immunology and Infectious Disease, Cumming School of Medicine, University of Calgary, Calgary, AB T2N 4N1, Canada.

[e]Li Ka Shing Institute of Virology and Discovery Lab, University of Alberta, Edmonton, AB T6G 2E1, Canada.

Equal author contribution

***Corresponding author:** Trushar R. Patel (email: trushar.patel@uleth.ca).

6.1 Abstract

Zika virus (ZIKV) infection remains a worldwide concern, and currently, no effective treatments or vaccines are available. Novel therapeutics are an avenue of interest that could probe viral RNA-human protein communication to stop viral replication. One specific RNA structure, G-quadruplexes (G4s), possess various roles in viruses and all domains of life, including transcription and translation regulation, and genome stability. Previous G4 studies on ZIKV using a **Q**uadruplex forming **G**-**R**ich **S**equences (QGRS) Mapper located a potential G-quadruplex sequence (PQS) in the 3' terminal region (TR) and was validated structurally using a 25-mer oligo. It is currently unknown if this structure is conserved and maintained in a large ZIKV RNA transcript, and its specific roles in viral replication. Using bioinformatic analysis and biochemical assays, we demonstrate that the ZIKV 3' TR G4 is conserved across all ZIKV isolates and maintains its structure in a 3' TR full-length transcript. We further established the G4 formation using Pyridostatin (PDS) and the BG4 G4-recognizing antibody binding assays. Our work also demonstrates that the human DEAD-box helicases, $DDX3X_{132-607}$ and $DDX17_{135-555}$, bind to the 3' TR, and that $DDX17_{135-555}$ unfolds the G4 present in the 3' TR. These findings provide a path forward in potential therapeutic targeting of DDX3X or DDX17’s binding to the 3' TR G4 region for novel treatments against ZIKV.

6.2 Introduction

ZIKV is an endemic, neurovirulent arbovirus whose primary vectors are *Aedes aegypti* and *Aedes albopictus* mosquitoes. Its primary transmission mode is the horizontal transfer of mosquito-infectious saliva during blood feeding [1]. Currently, the CDC has designated 86 countries as ZIKV risk areas [2]. ZIKV is popularly known for the 2015/2016 epidemic in Brazil; when during that time, it was revealed that the virus can be vertically transmitted from an infected pregnant woman to their fetus, which can lead to congenital disabilities such as microcephaly *in utero* [3]. ZIKV can also cause debilitating symptoms in adults, including Guillain-Barré Syndrome, fever, rash, headaches, and joint pain, and can also be sexually transmitted [3-5]. Currently, there are no treatments or vaccines available [6]. ZIKV possesses a positive-sense, non-segmented RNA genome of ~10.4 kilobases which is comprised of a 5' terminal region (TR) (~108 nt), 3 structural genes - capsid (C), membrane (prM), and envelope (E), 7 non-structural genes - NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5, followed by a long, 3' TR (~432 nt). The TRs play a vital role in viral genome cyclization, which regulates viral transcription and translation, alters host mRNA turnover, suppresses interferon response, and modulates the host-cell environment [7-13].

An area of interest for antiviral targeting are G-quadruplexes (G4s) - non-canonical structures identified in all domains of life, and almost all viral groups in the Baltimore classification [14]. Viral G4s are known to play critical roles throughout regulatory processes in their life cycles, such as transcriptional and translational activities, as well as alternative splicing [15-18]. These guanine-rich structures consist of at least 2 consecutive guanine tetrads Hoogsteen base pairing with each other in both DNA and RNA [19-21], altogether stabilized by a monovalent cation (preferably K$^+$) at the coordinating site (Fig. 6.1A). DNA G4s can adopt parallel, antiparallel, and hybrid conformations, whereas RNA G4s exclusively adopt propeller-type parallel topology due to the 2'-OH preventing syn-conformation torsion angles [22] (Fig. 6.1A).

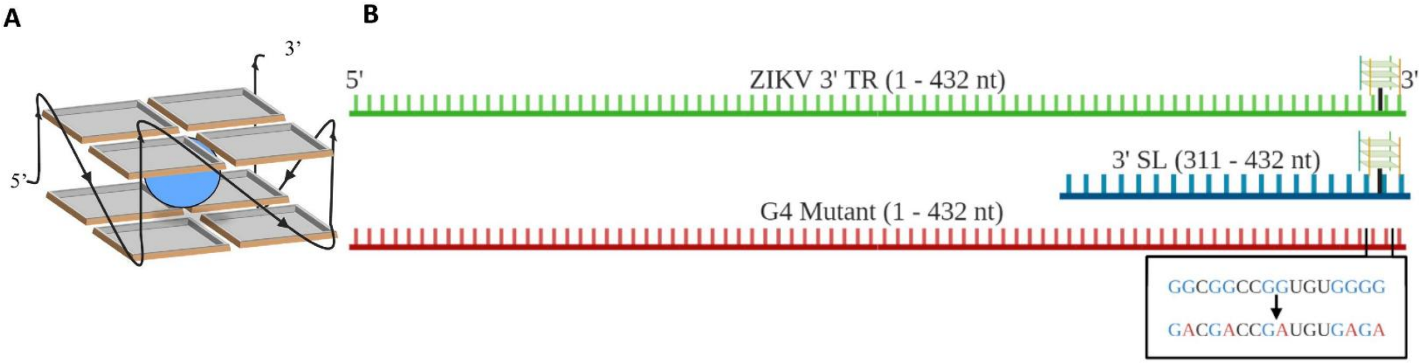


Figure 6.1 Schematic of a G-quadruplex and RNA transcripts under study. (A) represents a schematic of a G4 in parallel conformation with monovalent cations coordinated in the core of the stacked guanines. (B) Illustrates the transcripts that were used in the experiments, with the potential G-quadruplex sequence (PQS) at the ends of the 3' TR (green) and 3' SL (blue). The G4 mutant is shown in red, along with the G4 mutation sequence shown below compared to the wild-type sequence (created with BioRender.com).

It has been shown that many human proteins, including DExD-box helicases interact with G4s [23-25]. The DExD-box family proteins are involved in a wide range of functions, from embryonic development, cell proliferation, hematopoiesis, metabolism, immune response, cancer pathogenesis, inflammation, autoimmune diseases, and influencing of viral replication [26, 27]. We have previously demonstrated that the DEAD-box helicases, DDX3X$_{132\text{-}607}$ or DDX17$_{135\text{-}555}$, interact and unfold the intergenic region (IGR) and 5' non-coding region (NCR) of the Rift Valley Fever virus, and the Japanese encephalitis and Zika virus 5' TRs, respectively [28, 29]. DDX3X has been reported as a critical factor required for Dengue and Hepatitis C viral replication and as an inhibitory factor for West Nile virus replication [30, 31]. It has also been shown to interact with RNA G4s [32, 33]. Conversely, it is unclear if DDX17 directly binds to G4s, as it is involved in G4 binding with other proteins, but no direct interaction has been demonstrated [34, 35]. A previous study on the identification of ZIKV genomic RNA structure of both the African and Asian/American lineages using selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) identified a conserved, canonical stem-loop structure within the 3' terminal region (TR) (classified as the 3' stem-loop (SL)) [36]. Interestingly, a study by Fleming *et al*. (2016) found potential G-quadruplex sequences (PQS) within ZIKV using the Quadruplex forming G-Rich Sequences (QGRS) Mapper, and they identified and characterized a G4 contained in the 3' SL, albeit, using a 25-mer oligo short-range interaction approach [37, 38]. From the perspective of long-range RNA interactions, it is still unclear if a G4 is maintained in the 3' SL. Additionally, a previous study indicated that conformational switching could occur between G4 and canonical RNA structures, as in this study's case, a hairpin [39]. These events are based on the proximity of specific mono and divalent cations, which are altogether generated by RNA liquid-liquid phase separation [40]. From these recent discoveries, the 3' SL could exist as canonical and non-canonical states based on ion proximity.

Because of the lack of treatment options for ZIKV, investigation of alternative targets within the viral structure or lifecycle are being explored, G4s being one of them [41]. The G4 binding molecules, Braco-19, TMPyP4, Berberine, NiL, 360A, and PDS, have all been shown to interact with ZIKV G4s *in vitro* and cell culturing assays, and showed that these molecules altogether reduce viral replication and protein production to varying degrees [41, 42]. It remains unclear if the 3' TR G4 exists in a larger RNA structure containing long-range interactions, as previous studies focus on a short-range approach using short oligos. Moreover, deciphering the various human proteins interacting with this G4 could provide insight into its function during viral replication.

In this study, we investigated the presence of a G4 in the full-length ZIKV 3' TR. For this purpose, we first used bioinformatics analyses to demonstrate that the 3' SL G4 sequence is present in multiple ZIKV isolates and by implementing a long-range interaction approach with long RNA transcripts of the 3' TR (432 nt) and 3' SL (104 nt) (Fig. 6.1B), we were able to identify the presence of a G4 in the 3' TR transcript using the BG4 antibody, TMPyP4, and PDS, which all specifically bind to G4s. Subsequently, we elucidated that human DDX3X$_{132\text{-}607}$ and DDX17$_{135\text{-}555}$ interact with the 3' TR and that DDX17$_{135\text{-}555}$ unwinds the G4 present in the 3' SL in an ATP-dependent manner.

6.3 Materials and Methods

6.3.1 Bioinformatics analysis of conserved G4s in the 3' TR

ZIKV viral genome data was downloaded from NCBI Genbank on 14 January 2021, comprising 1028 ZIKV isolates. Filtering for isolates that cover the complete G4-containing 3' SL regions was performed with an Infernal covariance model [43] constructed from the 3' SL seed alignment of the Spondweni group featured in the viRNA GitHub repository (https://github.com/mtw/viRNA), resulting in 113 ZIKV isolates. Filtering these for a non-redundant set of 3' SL regions resulted in a total of 14 representative ZIKV isolates. The ZIKV isolate from Haiti used in this study (KU509998.3) is identical to strain MN577544.1, featured in our non-redundant set, throughout the entire 3' SL region. A structural multiple sequence alignment of the 3' SL region of 14 representative ZIKV isolates was computed with LocARNA [44], and visualized with Jalview [45].

6.3.2 RNA transcript expression of 3' TR, 3' SL, and G4 mutant

The ZIKV 3' TR (10375-10807) (432 nt) sequence was derived from ZIKV isolate from Haiti (GenBank: KU509998.3) [46]. This isolate also has a near identical sequence to the Asian/American lineage, one of the predominant lineages across the globe [47]. We designed 3 plasmid constructs from this sequence containing the full 3' TR, the 3' SL and a 3' TR with the G4 sequence mutated (G4 mutant). The G4 mutant construct was designed wherein 5 nucleotide mutations were made (5'-GGCGGCCGGUGUGGGG-3' → 5'-GACGACCGAUGUGAGA-3') (Fig. 6.1B) in the hypothesized G4 sequence in order to disrupt the G4 formation. All constructs were made using pUC57-Kan plasmids with a T7 promoter upstream of the desired transcript, with an *Xba*I cut site at the end of the transcript to linearize the plasmid such that T7 polymerase dissociates during *in vitro* transcription (IVT) reactions. The 3' TR and 3' SL plasmid constructs were commercially synthesized by Integrated DNA Technologies™, and BioBasic™ synthesized the G4 mutant.

The synthesized plasmids were transformed into competent *E. coli* NEB5α cells (New England Biolabs™), and the cells were further propagated to replicate the plasmid DNA for IVTs. The plasmid DNA was extracted, linearized using *Xba*I restriction digestion, and subsequently used for IVT reaction. The T7 polymerase-based (expressed and purified in-house) IVT reaction was performed by mixing transcription buffer (200 mM Tris-Cl (pH 7.5), 75 mM $MgCl_2$, 10 mM spermadine, and 50 mM NaCl), 100 mM DTT, 25 mM NTPs, 100 mM GMP, 0.5U/µL IPPase, 10 µM T7 polymerase, Ribolock, and the digested plasmid, and performed for 3 hours at 37°C. This was followed immediately by purification via size-exclusion chromatography (SEC) using a Cytiva Superdex™ 200 Increase 10/300 GL for the ZIKV 3' SL and a Sephacryl™ S-400 HR for the 3' TR and

G4 mutant RNA. The columns were pre-equilibrated in G4 buffer (10 mM HEPES (pH 7.5) and 100 mM KCl). For purity analysis, the peak fractions were analyzed on a 2% agarose gel. The purified RNA was then ethanol precipitated at -80°C in preparation for RNA labelling for subsequent experiments.

6.3.3 Expression and purification of BG4 antibody, as well as DDX17$_{135\text{-}555}$, and DDX3$_{132\text{-}607}$ proteins

The plasmid coding for the BG4 antibody was obtained from Addgene™ (pSANG10-3F-BG4) and transformed into *E. coli* BL21 DE3 for protein expression. The transformed cells were grown at 37°C until the optical density reached 0.6 at 600 nm. Subsequently, cells were induced with 1 mM isopropylthio-β-galactoside (IPTG) and were grown for 16 hours at 16°C, followed by cell harvesting through centrifugation. The harvested cells were resuspended in lysis buffer (50 mM Tris-Cl (pH 7.5), 500 mM NaCl, 0.2% tween, 5% glycerol, 5 mM β-mercaptoethanol, 10 mM imidazole), supplemented with 0.1 mg/mL lysozyme, 5,200 units of DNase I (Thermofisher™), 2 mM PMSF, as well as protease inhibitors (pepstatin, aprotinin, and AEBSF) from Biobasic™. The cell suspension was sonicated and clarified using centrifugation (30,000 x g); the clarified lysate was loaded onto a 1 mL Histrap™ high-performance column from Cytiva™ mounted on an Äkta Start™ system and purified using an elution gradient from 10 mM to 150 mM imidazole supplemented to the lysis buffer. The fractions were purified via SEC using a Superdex™ 75 increase 10/300 GL (Cytiva™) pre-equilibrated with 1x PBS mounted on an Äkta Pure™ system. The fractions were analyzed on a 12% SDS-PAGE gel pooled and concentrated to 14 µM and stored at -20°C until further use.

DDX3X$_{132-607}$, DDX17$_{135-555}$, and DDX17 mutant are altogether truncated versions of the protein, and each were expressed according to protocols described previously using an *E. coli* expression system similar to the BG4 antibody [28, 29]. DDX17 mutant was also purified using a pET28a cloned vector with glycine mutations on amino acids which coordinate ATP hydrolysis; the Q motif, which interacts with the adenine of ATP, as well as motifs 1 & 2 in the ATP binding domain which interacts with the triphosphate [48-50]. The mutant assays were performed to prevent ATP hydrolysis, and consequently, helicase activity.

6.3.4 Alexa Fluor™ 488 labelling of the RNA

Each purified RNA transcript was labeled at the 5' end with Alex fluor™ 488. 1.25 mg of 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride was mixed with 8 µL of concentrated RNA in water and 5 µL Alexa Fluor™ 488 in 0.2 M KCl, and 20 µL of 0.1 M imidazole (pH 6). All reactions were incubated at room temperature for 18 hours. The RNA was then diluted in 460 µL of G4 buffer (10 mM HEPES (pH 7.5) and 100 mM KCl) and loaded onto a Cytiva™ Superdex™ 200 Increase 10/300 GL in order to purify the labeled RNA from the unlabeled fluorophore. Each RNA sample was then collected and heat-cooled at 95℃ for 5 minutes, followed by room temperature incubation for 15 minutes. The RNA was tested for fluorescence counts on the NanoTemper Technologies Monolith® NT.115 MST device to ensure optimal labelling for downstream experiments, similar to previous experiments that used Alex Fluor™ 488 for labelling their nucleic acids [51-53]. The labeled RNA was then vitrified in liquid nitrogen and stored at -80℃ until downstream experimentation.

6.3.5 Interaction studies using microscale thermophoresis.

A Monolith® NT.115 MST from NanoTemper Technologies was used to evaluate the affinity of RNA constructs with their ligands. The ZIKV 3' TR, G4 mutant, and 3' SL RNA transcripts were evaluated at 80 nM at 100% excitation at medium MST power for the BG4 binding affinity, and PDS binding check runs. For the BG4 runs, the samples were serially diluted two-fold in the MST capillaries concentrations ranging from BG4 from 6.8 µM – 0.8 nM, DDX3X$_{132\text{-}607}$ from 7 µM – 14.6 nM, and DDX17$_{135\text{-}555}$ from 19 µM – 0.58 nM. All MST data were collected in triplicate at room temperature using the Nano-Blue filter and medium IR-laser power for BG4, while high power was used for DDX3X$_{132\text{-}607}$ or DDX17$_{135\text{-}555}$. The dissociation constants were calculated using the Monolith NT.115 analysis software by plotting the ligand percent bound, and for the PDS runs, the average response amplitude was recorded in triplicate, and the Monolith NT.115 analysis software provided the signal-to-noise ratio values.

6.3.6 Unwinding assays of the ZIKV 3' TR G4

To assess DDX3X$_{132\text{-}607}$ and DDX17$_{135\text{-}555}$'s ability to unwind the non-canonical structured area in the 3' SL, we used MST and a complementary RNA (5'-AGUUUCCACCACGCUGGCCGCCA-3') that will only base pair to the G4 portion of the target RNA, if it is unfolded by DDX3X$_{132\text{-}607}$ or DDX17$_{135\text{-}555}$, as performed previously [28, 29]. When fluorescence and/or migratory differences occur between the BSA control and the helicase, this provides evidence that the helicase unfolded the non-canonical region, as the oligo base pairing is causing a change in fluorescence, shape, charge, and/or hydration shell of the target RNA [54]. The reaction was performed using 1 µM ZIKV 3' SL RNA sample, 4.25 µM ATP, 40 nM Cy-5- DNA Oligo and 20 µM of either DDX3X$_{132\text{-}}$

$_{607}$ or DDX17$_{135\text{-}555}$. BSA was selected as a negative control. The reaction was incubated for 15 minutes prior to the measurements. The experiments were performed 8 times in total over two binding check runs on the Nanotemper Monolith® NT.115. at 50% excitation and medium MST power, the signal-to-noise ratio was assessed for each run to determine successful unwinding activity, which, again, a signal-to-noise ratio of ≥5 indicates that binding is occurring.

A spectrofluorimetric unwinding assay was also performed using a Quanta Master 60 fluorescence spectrometer (Photon Technology International). The reactions were performed using a mixture of 2 µM protein, 50 nM RNA, and 3.5 µM ThT - a well characterized G4 binding biosensor [55]. The mixture was incubated for 5 minutes to allow ThT to bind to the G4 on the RNA, followed by a fluorescence measurement taken at 425 nm excitation and emission collected from 450 – 510 nm with a step size of 1 nm increments. Following collection, 60 µM of ATP was titrated into the mixture, followed by a 5-minute incubation to allow the helicase to unfold the G4, preventing ThT from binding, as the G4s is unfolded, and the same fluorescence parameters were used to collect the emission. The change in fluorescence was then plotted (n=3), and a t-test was performed to check for statistical significance. DDX17$_{135\text{-}555}$, DDX3X$_{132\text{-}607}$, and a control, BSA, were all used to validate the MST results of unwinding.

6.4 Results

6.4.1 The G4 sequence contained in the 3' SL of the 3' TR is conserved across the globe in ZIKV isolates.

To obtain a comprehensive view of the sequence diversity within the 3' TR of all known ZIKV isolates, we collected all complete ZIKV genomes from NCBI Genbank [56] and extracted the genomic regions that fold into the 3' SL element. Filtering for non-redundant sequences in the 3' SL region resulted in 14 representative ZIKV isolates, comprising both African and Asian/American ZIKV lineages. A structural multiple sequence alignment is shown in Fig. 6.2, highlighting strong primary sequence conservation of the region of interest that could potentially form a G4 structure.

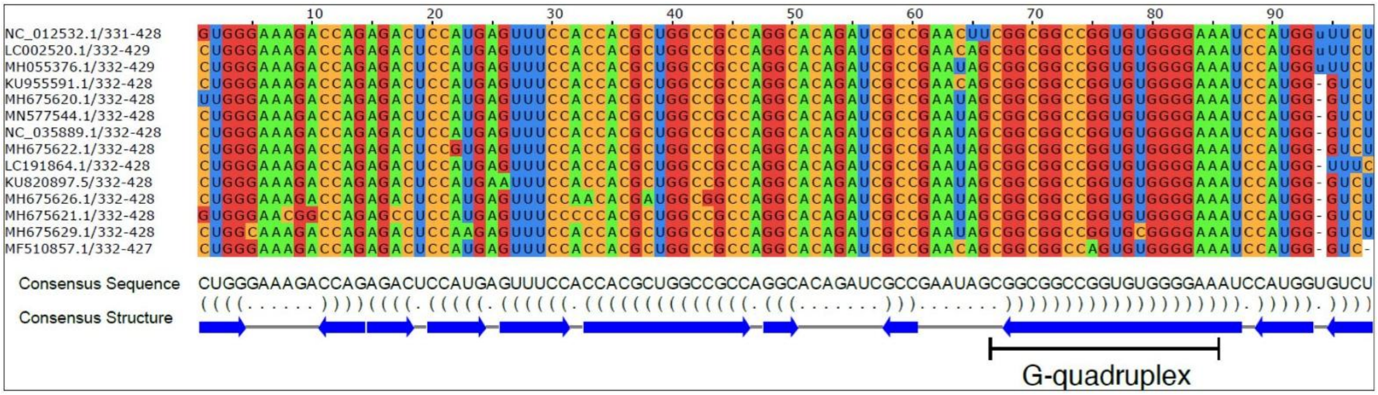


Figure 6.2 The 3' SL G4 sequence is conserved across 1028 ZIKV isolates. 14 non-redundant ZIKV 3' SL sequences from African and Asian/American lineages were aligned using the LocARNA software [44]. The consensus sequence and its accompanying dot-bracket notation are also included, noting that the software predicts a canonical structure within the dot-bracket notation.

6.4.2 ZIKV 3' TR contains a G4 structure.

We performed microscale thermophoresis (MST) binding checks with PDS, and binding affinity studies with the BG4 antibody, altogether with the 3' TR of ZIKV to investigate if the 3' TR contains a G4. PDS is a small molecule that selectively binds to the top quartet of a G4 and stabilizes the non-canonical structure [57-60]. PDS showed a significant response amplitude in the 3' TR and 3' SL compared to the G4 mutant in samples treated with PDS vs. no PDS (Fig. 6.3A). These results suggest that any potential for G4 formation in the G4 mutant is significantly impeded or non-existent, as PDS is below the signal-to-noise (SNR) threshold, and the response amplitude (RA) is significantly lower than the 3' TR and 3' SL. The Nanotemper™ analysis software states that a binding event evidently occurs when the SNR value is ≥ 5 when comparing treated vs. untreated samples. Previously, it was demonstrated that a ~20 nt oligo from the 3' SL of ZIKV contains a G4 structure using PDS [60] indicating that PDS is a reliable molecule in the study/identification of G4s.

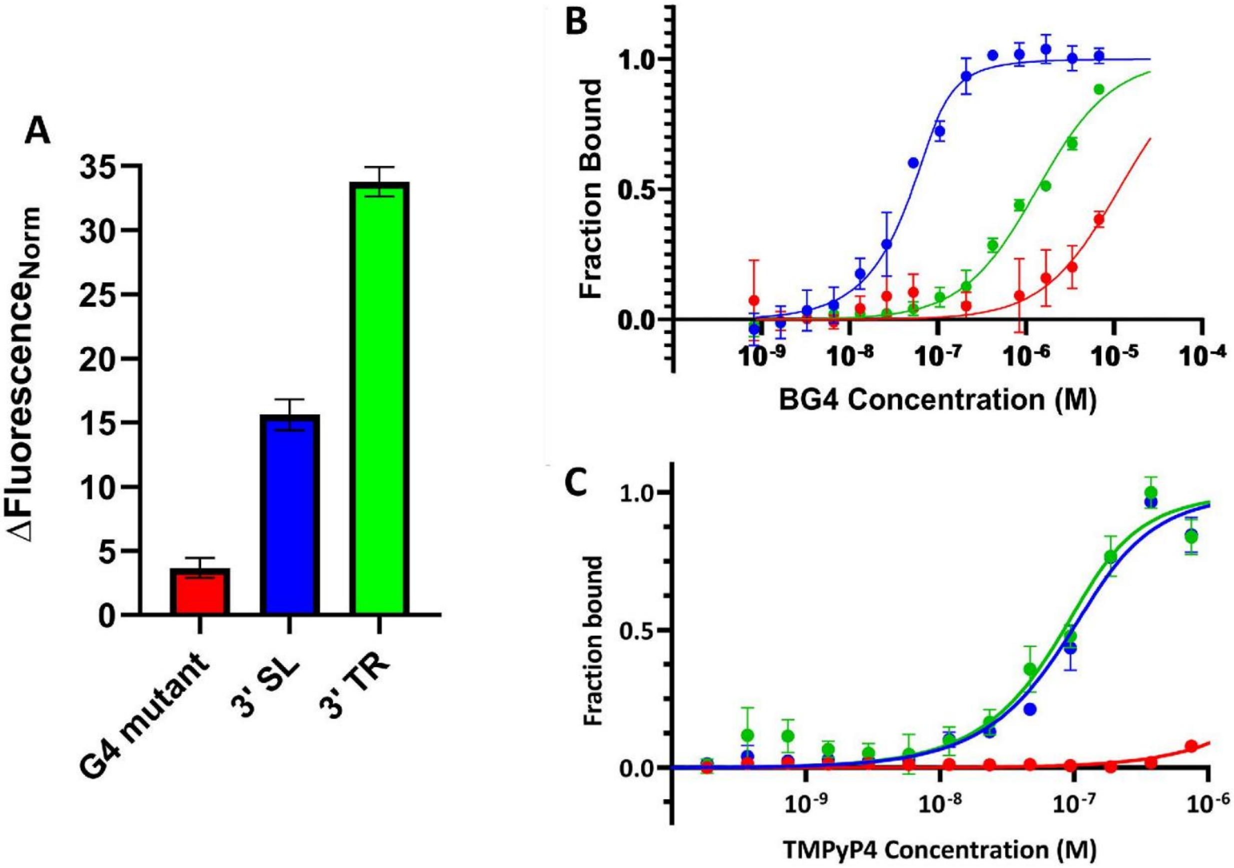


Figure 6.3 A G4 exists in the 3' SL of the 3' TR. (**A**) represents the change in normalized fluorescence data with RNA alone compared to RNA + PDS (n=3). (**B** & **C**) demonstrate binding affinities of BG4 and TMPyP4 with ZIKV 3' SL (Blue), 3' TR (Green), and 3' G4 mutant (Red). The determined K_d for each transcript for BG4 were: 3' SL = 0.0106 ± 0.0023 µM, 3' TR = 1.27 ± 0.087 µM, 3' G4 mutant = 11.11 ± 1.71 µM, and for TMPyP4: 3' SL = 0.0471 ± 0.0079 µM, 3' TR = 0.0339 ± 0.0117 µM, 3' G4 mutant = no binding (n=3).

We further investigated the presence of a G4 using an antibody (BG4) that specifically targets the G4 structure. The BG4 antibody is a well-established, reliable antibody that has been widely used to identify G4s of both DNA and RNA *in vivo* and *in vitro* [61-71]. The MST experiments using the BG4 antibody suggested that BG4 interacts with the 3' SL at the nanomolar range, 0.0106 ± 0.0023 µM (Fig. 6.3B, blue line), whereas the binding affinity of the 3' TR was determined to be 1.27 ± 0.08 µM (Fig. 6.3B, green line). The binding studies of BG4 with G4 mutant suggested a K_d of 11.11 ± 1.71 µM (Fig. 6.3B, red line), which is ten-fold weaker compared to the wild-type sequence. This change in the K_d value suggests that the mutant sequence cannot form a G4. A potential reason for the kinetic differences between the 3' SL and 3' TR could be due to steric/structural hindrance specific to the ZIKV 3' TR that is not present in the 3' SL, as previous literature shows that kinetics can differ depending on access to the binding site [72]. This data was further supported by performing binding affinity MST studies with TMPyP4 (Fig. 6.3C), which is also a well characterized G4 binding/unfolding small molecule [73-75]. The affinity determined for each RNA with TMPyP4 was: 3' SL = 0.0471 ± 0.0079 µM, 3' TR = 0.0339 ± 0.0117 µM, 3' G4 mutant = no binding.

6.4.3 DDX3X$_{132\text{-}607}$ and DDX17$_{135\text{-}555}$ interact with the ZIKV 3' TR

The interaction studies between the ZIKV 3' TR and the human helicases, DDX3X$_{132\text{-}607}$ and DDX17$_{135\text{-}555}$, were performed using MST. Fig. 6.4 shows the binding affinities of DDX17$_{135\text{-}555}$ and DDX3X$_{132\text{-}607}$ with the ZIKV 3' TR, with a K_d of 1.16 ± 0.02 µM and 3.70 ± 0.10 µM, respectively. Additionally, these interactions were confirmed through a native-PAGE and agarose Electrophoretic Mobility Shift Assay (EMSA) where a shift is observed in concentrations measured around the K_d for both DDX3X$_{132\text{-}607}$ and DDX17$_{135\text{-}555}$. These affinities align with previous studies on the ZIKV 5' TR, which suggested that DDX3X$_{132\text{-}607}$ binds with an affinity at 7.05 µM [29]. It is currently unclear what role(s) these helicases play in viral replication, as it could be detrimental or advantageous to the viral life cycle [27, 76].

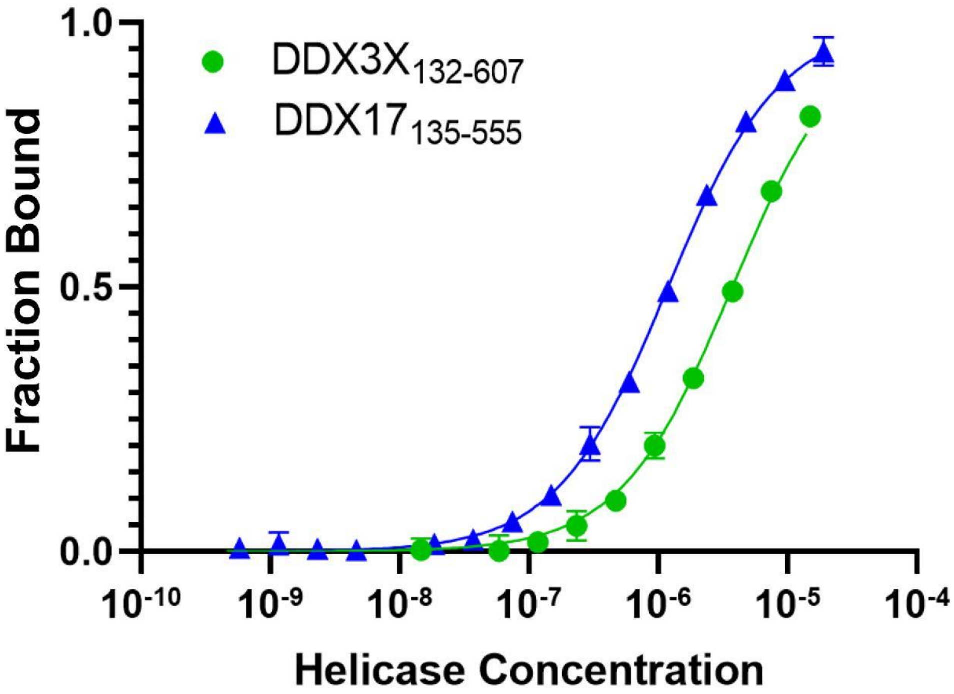


Figure 6.4 $DDX3X_{132-607}$ and $DDX17_{135-555}$ bind to the 3' TR. MST traces demonstrating the interaction of the 3' TR of ZIKV with $DDX3X_{132-607}$ (green) and $DDX17_{135-555}$ (blue). RNA concentration was a constant 50 nM, while the dilution series started from 19 µM for $DDX17_{135-555}$ and 15 µM for $DDX3X_{132-607}$ (n=3). Measurements were performed using 20% excitation and high MST power. The K_d was determined to be 1.16 ± 0.02 µM for $DDX17_{135-555}$ and 3.70 ± 0.10 µM for $DDX3X_{132-607}$.

6.4.4 $DDX17_{135-555}$ unwinds the 3' TR G4 using the 3' SL

Once we established that both helicases interact with ZIKV 3' TR, we asked if $DDX17_{135-555}$ and $DDX3X_{132-607}$ can unfold the G4 in the 3' SL, as there is currently no evidence demonstrating that DDX3X or DDX17 unfolds G4s. The helicase assays were performed using the 3' SL to investigate if $DDX3X_{132-607}$ and $DDX17_{135-555}$ unwind the G4 present in 3' SL. MST helicase assay utilizes a fluorescently labeled oligo that was designed to be complementary to a portion of the target RNA that was predicted to be double-stranded, as previously described [29]. If the presence of the oligo results in a change in the migration of the fluorescently labeled molecules, we can infer that the RNA was unwound, permitting the binding to occur. The MST experiments indicated that only DDX17$_{135-555}$ caused a shift in the fluorescent migration as compared to the BSA control and DDX17 mutant (Fig. 6.5A). An unpaired *t*-test was performed for DDX3X$_{132-607}$, DDX17$_{135-555}$, and DDX17 mutant, and the *p*-value at 95% confidence indicated a value of 0.029 for DDX3X$_{132-607}$, 0.001 for DDX17$_{135-555}$, and no statistical significance for DDX17 mutant. Collectively, these experiments suggest that the Cy5-DNA oligo binds to the unwound RNA in the presence of DDX17$_{135-555}$. The signal-to-noise (SNR) ratio for DDX17$_{135-555}$ were all above 5, which confirms the assay was successful.

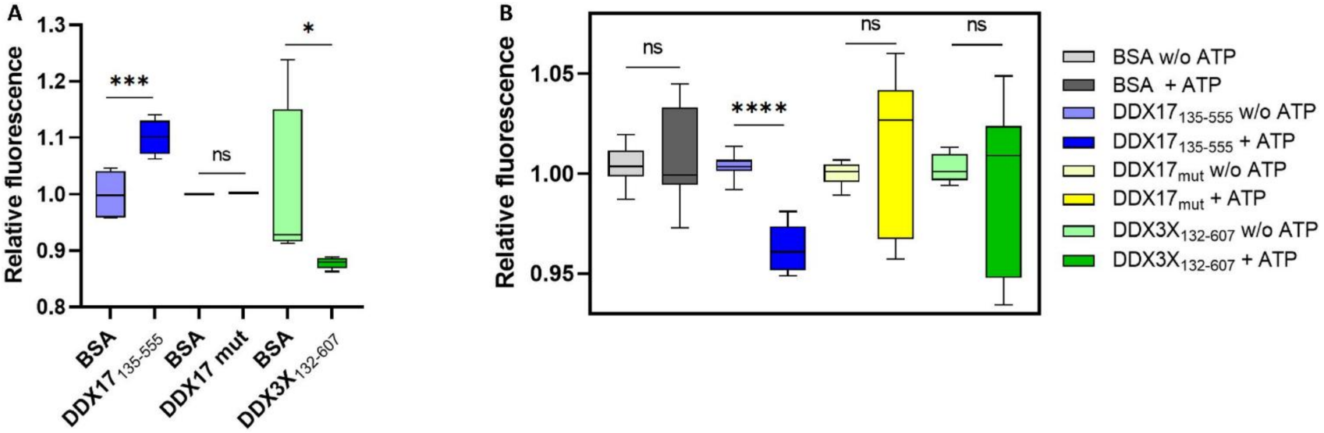


Figure 6.5 DDX17$_{135-555}$ unfolds the 3' SL G4. A & **B** are the helicase assay experiment performed using Microscale thermophoresis (**A**) and fluorescence spectroscopy (**B**). Experiments in **A** contain the complementary labeled oligonucleotide that base pairs to the G4 sequence if unfolded by the helicase protein in the presence of ATP with the relative fluorescence plotted (n=4). During this binding check, it is essential to run a BSA control alongside each individual sample being tested, as BSA is the reference point for each individual run. The significance of these runs is the SNR in the fluorescence difference. The fluorescence spectroscopy in **B** contain the G4 binding biosensor, ThT with the relative fluorescence differences measured between the presence and absence of ATP in the reaction mixture (n=3). The decrease in fluorescence in **B** indicates that the G4 is unfolding in the 3' SL, hence, ThT is unable to bind. Like the MST analysis in **A**, the fluorescence in **B** need not be normalized, as a comparison of the highest fluorescence peaks within each measurement is collected. When comparing the fluorescence, it was performed in pairs, each with a BSA control for the individual proteins, rather than a comparison of the different proteins against each other. In **A**, statistically significant shifts in each assay indicate that the oligo is bound to the 3' SL. A T-test was performed for all assays to signify

statistical relevance of our data. The asterisks indicate the p-value (ns = $P>0.05$, * = $P\leq0.05$, ** = $P\leq0.01$, *** = $P\leq0.001$, **** = $P\leq0.0001$).

We also performed a spectrofluorimetric helicase assay by replacing the complementary oligo to ThT, a G4-binding fluorescence sensor that stacks to G-tetrads stabilizing K^+ ions [77]. This system allows for the collection of the ThT fluorescence emission when it recognizes a properly folded G4 compared to unwound molecules in the presence of helicases and ATP. A fluorescence decrease was observed upon the addition of ATP into $DDX17_{135-555}$ and RNA sample, whereas for BSA, $DDX3X_{132-607}$, and DDX17 mutant showed no statistically significant change (Fig. 6.5B). These results suggest that $DDX17_{135-555}$ unwound the G4, specifically on the 3' SL preventing ThT from binding/fluorescing, whereas $DDX3X_{132-607}$, and DDX17 mutant are unable to unwind the same RNA. $DDX3X_{132-607}$ and DDX17 mutant also failed to produce a significant change in the fluorescent migration of the oligo in the MST assays, which is seen in the SNR ratio, which was below 5. These combined results demonstrate a previously unreported function of $DDX17_{135-555}$: unwinding of a G4. The MST studies also demonstrate that while $DDX17_{135-555}$ can unfold the 3' SL G4, $DDX3X_{132-607}$ does not.

6.5 Discussion

Our work demonstrates the presence of G4 in ZIKV 3' TR and establishes that human helicase $DDX17_{135-555}$ can unwind the G4 structure. These results unveil a previously unreported function of DDX17. Further work is required to elucidate the function(s) of the 3' TR G4 in ZIKV life cycle. DDX3X has been extensively studied for its ability to interact with various RNA structures, including G4 [78]. Consistent with this finding, our previous study demonstrated the capacity of DDX3X$_{132\text{-}607}$ unwinding secondary structure within the ZIKV and Japanese Encephalitis virus 5' TRs [29]. We hypothesize that the observed inability to unwind G4 structures might be attributed to a structural feature within the substrate rather than the absence of a functional domain within the helicase. However, it is still unclear whether DDX3X can unwind G4s [79, 80]. Previous work has led to interesting hypotheses such as the role of G4s in stabilizing the 3' TR of the viral genome, serving as a ‘capped’ degradation-resistant region comparable to G overhangs in telomeres, evasion of the host intracellular immune response, assisting in viral transcription initiation, or act as another element in host protein ‘sponging’, as this structure is contained in sub-flaviviral RNA [9, 81-85]. Moreover, protein sponging could also be assisted with a G4 serving as a nucleation point for RNA phase separation, potentially giving the virus more control of ion proximity and structural switching between a G4 and stem-loop [39, 40]. This would potentially provide ZIKV an extension of the number of human proteins that can be recruited to the terminal regions as well, thus, providing a more streamlined replication. Furthermore, it has also been shown that 12 regions of the 3' TR contain N6-methyladenosines (m^6A) enrichment, and 8 out of the 12 m^6A regions were contained on the 3' SL G4. Previous literature showed that in the case of ZIKV, when m^6A was suppressed, it increased viral replication, suggesting that m^6A could be a mechanism the cell uses to slow viral replication [86]. Overall, further investigation into the role of G4 from ZIKV 3' TR is required.

ZIKV remains a significant public health concern globally, and due to concerns of potential antibody-dependent enhancement caused by antibodies generated from flaviviral vaccines or infection, novel treatment avenues to stop viral replication are urgently required. To this end, our work demonstrated that ZIKV contains a G4 within the 3' SL of the 3' TR that is maintained across multiple ZIKV isolates and is unfolded by human DDX17$_{135\text{-}555}$. This discovery allows for further investigation to exploit this structured region by providing potential treatments using G4 binding molecules or other molecules that disrupt human DDX17 – 3' SL interaction. Further studies of the function of the 3' SL G4 during viral replication are essential, but the foundation of identification of the G4 and two RNA-binding proteins allows for a benchmark to start exploring this region as a target to attempt to stop ZIKV replication, providing a novel treatment for ZIKV infection.

6.6 References

1. Musso, D. and D.J. Gubler, *Zika Virus.* Clin Microbiol Rev, 2016. **29**(3): p. 487-524.
2. Prevention, C.o.D.C.a. *Zika Travel Information*. 2022; Available from: https://wwwnc.cdc.gov/travel/page/zika-information.
3. Tang, B.L., *Zika virus as a causative agent for primary microencephaly: the evidence so far.* Archives of Microbiology, 2016. **198**(7): p. 595-601.
4. Barzon, L., et al., *Infection dynamics in a traveller with persistent shedding of Zika virus RNA in semen for six months after returning from Haiti to Italy, January 2016.* Euro Surveill, 2016. **21**(32).
5. Mansuy, J.M., et al., *Zika virus: high infectious viral load in semen, a new sexually transmitted pathogen?* Lancet Infect Dis, 2016. **16**(4): p. 405.
6. Administration, T.U.S.F.a.D. *Zika Virus Response Updates from FDA*. 2022; Available from: https://www.fda.gov/emergency-preparedness-and-response/mcm-issues/zika-virus-response-updates-fda.
7. Sirohi, D. and R.J. Kuhn, *Zika Virus Structure, Maturation, and Receptors.* J Infect Dis, 2017. **216**(suppl_10): p. S935-s944.
8. Donald, C.L., et al., *Full Genome Sequence and sfRNA Interferon Antagonist Activity of Zika Virus from Recife, Brazil.* PLOS Neglected Tropical Diseases, 2016. **10**(10): p. e0005048.
9. Michalski, D., et al., *Zika virus noncoding sfRNAs sequester multiple host-derived RNA-binding proteins and modulate mRNA decay and splicing during infection.* J Biol Chem, 2019. **294**(44): p. 16282-16296.
10. Sanford, T.J., et al., *Circularization of flavivirus genomic RNA inhibits de novo translation initiation.* Nucleic Acids Res, 2019. **47**(18): p. 9789-9802.
11. Wolfinger, M.T., *Functional RNA Structures in the 3'UTR of Mosquito-Borne Flaviviruses.* Virus Bioinformatics 2021: p. 65-100.
12. Chavali, P.L., et al., *Neurodevelopmental protein Musashi-1 interacts with the Zika genome and promotes viral replication.* Science, 2017. **357**(6346): p. 83-88.
13. Schneider, A.d.B. and M.T. Wolfinger, *Musashi binding elements in Zika and related Flavivirus 3'UTRs: A comparative study in silico.* Scientific Reports, 2019. **9**(1): p. 6911.
14. Ruggiero, E. and S.N. Richter, *Viral G-quadruplexes: New frontiers in virus pathogenesis and antiviral therapy.* Annu Rep Med Chem, 2020. **54**: p. 101-131.
15. Métifiot, M., et al., *G-quadruplexes in viruses: function and potential therapeutic applications.* Nucleic Acids Res, 2014. **42**(20): p. 12352-66.
16. Patel, P.K., N.S. Bhavesh, and R.V. Hosur, *NMR observation of a novel C-tetrad in the structure of the SV40 repeat sequence GGGCGG.* Biochem Biophys Res Commun, 2000. **270**(3): p. 967-71.
17. Meinke, G., et al., *Structural Based Analyses of the JC Virus T-Antigen F258L mutant Provides Evidence for DNA Dependent Conformational Changes in the C-Termini of Polyomavirus Origin Binding Domains.* PLoS Pathog, 2016. **12**(1): p. e1005362.
18. Murat, P., et al., *G-quadruplexes regulate Epstein-Barr virus–encoded nuclear antigen 1 mRNA translation.* Nature Chemical Biology, 2014. **10**(5): p. 358-364.
19. Burge, S., et al., *Quadruplex DNA: sequence, topology and structure.* Nucleic Acids Res, 2006. **34**(19): p. 5402-15.

20. Wang, Y. and D.J. Patel, *Solution structure of a parallel-stranded G-quadruplex DNA.* J Mol Biol, 1993. **234**(4): p. 1171-83.
21. Bhattacharyya, D., G. Mirihana Arachchilage, and S. Basu, *Metal Cations in G-quadruplex Folding and Stability.* Front Chem, 2016. **4**: p. 38.
22. Fay, M.M., S.M. Lyons, and P. Ivanov, *RNA G-quadruplexes in Biology: Principles and Molecular Mechanisms.* J Mol Biol, 2017. **429**(14): p. 2127-2147.
23. Linder, P., et al., *Birth of the D-E-A-D box.* Nature, 1989. **337**(6203): p. 121-122.
24. Zhang, X., et al., *Chemical profiling of DNA G-quadruplex-interacting proteins in live cells.* Nat Chem, 2021. **13**(7): p. 626-633.
25. Brázda, V., et al., *DNA and RNA quadruplex-binding proteins.* Int J Mol Sci, 2014. **15**(10): p. 17493-517.
26. Andrisani, O., et al., *Biological functions of DEAD/DEAH-box RNA helicases in health and disease.* Nature Immunology, 2022. **23**(3): p. 354-357.
27. Meier-Stephenson, V., et al., *DEAD-box helicases: the Yin and Yang roles in viral infections.* Biotechnol Genet Eng Rev, 2018. **34**(1): p. 3-32.
28. Nelson, C.R., et al., *Human DDX17 Unwinds Rift Valley Fever Virus Non-Coding RNAs.* Int J Mol Sci, 2020. **22**(1).
29. Nelson, C., et al., *Human DDX3X Unwinds Japanese Encephalitis and Zika Viral 5' Terminal Regions.* Int J Mol Sci, 2021. **22**(1).
30. Ariumi, Y., et al., *DDX3 DEAD-box RNA helicase is required for hepatitis C virus RNA replication.* J Virol, 2007. **81**(24): p. 13922-6.
31. Brai, A., et al., *DDX3X Helicase Inhibitors as a New Strategy To Fight the West Nile Virus Infection.* Journal of Medicinal Chemistry, 2019. **62**(5): p. 2333-2347.
32. Varshney, D., et al., *RNA G-quadruplex structures control ribosomal protein production.* Scientific Reports, 2021. **11**(1): p. 22735.
33. Herdy, B., et al., *Analysis of NRAS RNA G-quadruplex binding proteins reveals DDX3X as a novel interactor of cellular G-quadruplex containing transcripts.* Nucleic Acids Research, 2018. **46**(21): p. 11592-11604.
34. Fortuna, T.R., et al., *DDX17 is involved in DNA damage repair and modifies FUS toxicity in an RGG-domain dependent manner.* Acta Neuropathol, 2021. **142**(3): p. 515-536.
35. Dardenne, E., et al., *RNA helicases DDX5 and DDX17 dynamically orchestrate transcription, miRNA, and splicing programs in cell differentiation.* Cell Rep, 2014. **7**(6): p. 1900-13.
36. Li, P., et al., *Integrative Analysis of Zika Virus Genome RNA Structure Reveals Critical Determinants of Viral Infectivity.* Cell Host Microbe, 2018. **24**(6): p. 875-886.e5.
37. Kikin, O., L. D'Antonio, and P.S. Bagga, *QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences.* Nucleic Acids Res, 2006. **34**(Web Server issue): p. W676-82.
38. Fleming, A.M., et al., *Zika Virus Genomic RNA Possesses Conserved G-quadruplexes Characteristic of the Flaviviridae Family.* ACS Infectious Diseases, 2016. **2**(10): p. 674-681.
39. Bugaut, A., P. Murat, and S. Balasubramanian, *An RNA hairpin to G-quadruplex conformational transition.* J Am Chem Soc, 2012. **134**(49): p. 19953-6.
40. Liu, X., et al., *G-quadruplex-Induced Liquid-Liquid Phase Separation in Biomimetic Protocells.* J Am Chem Soc, 2021. **143**(29): p. 11036-11043.

41. Majee, P., et al., *Inhibition of Zika virus replication by G-quadruplex-binding ligands.* Mol Ther Nucleic Acids, 2021. **23**: p. 691-701.
42. Zou, M., et al., *G-quadruplex binder pyridostatin as an effective multi-target ZIKV inhibitor.* Int J Biol Macromol, 2021. **190**: p. 178-188.
43. Nawrocki, E.P. and S.R. Eddy, *Infernal 1.1: 100-fold faster RNA homology searches.* Bioinformatics, 2013. **29**(22): p. 2933-2935.
44. Will, S., et al., *Inferring Noncoding RNA Families and Classes by Means of Genome-Scale Structure-Based Clustering.* PLOS Computational Biology, 2007. **3**(4): p. e65.
45. Clamp, M., et al., *The Jalview Java alignment editor.* Bioinformatics, 2004. **20**(3): p. 426-427.
46. Lednicky, J., et al., *Zika Virus Outbreak in Haiti in 2014: Molecular and Clinical Data.* PLoS Negl Trop Dis, 2016. **10**(4): p. e0004687.
47. Beaver, J.T., et al., *Evolution of Two Major Zika Virus Lineages: Implications for Pathology, Immune Response, and Vaccine Development.* Front Immunol, 2018. **9**: p. 1640.
48. Ali, M.A.M., *The DEAD-box protein family of RNA helicases: sentinels for a myriad of cellular functions with emerging roles in tumorigenesis.* International Journal of Clinical Oncology, 2021. **26**(5): p. 795-825.
49. Ali, M.A.M., *DEAD-box RNA helicases: The driving forces behind RNA metabolism at the crossroad of viral replication and antiviral innate immunity.* Virus Research, 2021. **296**: p. 198352.
50. Tanner, N.K., et al., *The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis.* Mol Cell, 2003. **11**(1): p. 127-38.
51. Yoo, B.C., et al., *A systemic small RNA signaling system in plants.* Plant Cell, 2004. **16**(8): p. 1979-2000.
52. Johnson, R.F., et al., *Ebola virus VP35-VP40 interaction is sufficient for packaging 3E-5E minigenome RNA into virus-like particles.* J Virol, 2006. **80**(11): p. 5135-44.
53. Kynast, R.G., et al., *Dissecting the maize genome by using chromosome addition and radiation hybrid lines.* Proc Natl Acad Sci U S A, 2004. **101**(26): p. 9921-6.
54. Huang, L. and C. Zhang, *Microscale Thermophoresis (MST) to Detect the Interaction Between Purified Protein and Small Molecule.* Methods Mol Biol, 2021. **2213**: p. 187-193.
55. Xu, S., et al., *Thioflavin T as an efficient fluorescence sensor for selective recognition of RNA G-quadruplexes.* Scientific Reports, 2016. **6**(1): p. 24793.
56. Clark, K., et al., *GenBank.* Nucleic Acids Res, 2016. **44**(D1): p. D67-72.
57. Zhang, S., Y. Wu, and W. Zhang, *G-quadruplex structures and their interaction diversity with ligands.* ChemMedChem, 2014. **9**(5): p. 899-911.
58. Rodriguez, R., et al., *Small-molecule-induced DNA damage identifies alternative DNA structures in human genes.* Nat Chem Biol, 2012. **8**(3): p. 301-10.
59. Moruno-Manchon, J.F., et al., *The G-quadruplex DNA stabilizing drug pyridostatin promotes DNA damage and downregulates transcription of Brca1 in neurons.* Aging (Albany NY), 2017. **9**(9): p. 1957-1970.
60. Hou, Y., et al., *G-quadruplex inducer/stabilizer pyridostatin targets SUB1 to promote cytotoxicity of a transplatinum complex.* Nucleic Acids Research, 2022. **50**(6): p. 3070-3082.

61. Javadekar, S.M., et al., *Characterization of G-quadruplex antibody reveals differential specificity for G4 DNA forms.* DNA Res, 2020. **27**(5).
62. Komůrková, D., A. Svobodová Kovaříková, and E. Bártová, *G-quadruplex Structures Colocalize with Transcription Factories and Nuclear Speckles Surrounded by Acetylated and Dimethylated Histones H3.* Int J Mol Sci, 2021. **22**(4).
63. Biffi, G., et al., *Quantitative visualization of DNA G-quadruplex structures in human cells.* Nat Chem, 2013. **5**(3): p. 182-6.
64. Mao, S.Q., et al., *DNA G-quadruplex structures mold the DNA methylome.* Nat Struct Mol Biol, 2018. **25**(10): p. 951-957.
65. Xu, Y.Z., et al., *Activation-induced cytidine deaminase localizes to G-quadruplex motifs at mutation hotspots in lymphoma.* NAR Cancer, 2020. **2**(4): p. zcaa029.
66. Byrd, A.K., et al., *Evidence That G-quadruplex DNA Accumulates in the Cytoplasm and Participates in Stress Granule Assembly in Response to Oxidative Stress.* J Biol Chem, 2016. **291**(34): p. 18041-57.
67. David, A.P., et al., *CNBP controls transcription by unfolding DNA G-quadruplex structures.* Nucleic Acids Res, 2019. **47**(15): p. 7901-7913.
68. Biffi, G., et al., *Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells.* Nat Chem, 2014. **6**(1): p. 75-80.
69. Canesin, G., et al., *Scavenging of Labile Heme by Hemopexin Is a Key Checkpoint in Cancer Growth and Metastases.* Cell Rep, 2020. **32**(12): p. 108181.
70. Biffi, G., et al., *Elevated levels of G-quadruplex formation in human stomach and liver cancer tissues.* PLoS One, 2014. **9**(7): p. e102711.
71. Varshney, D., et al., *RNA G-quadruplex structures control ribosomal protein production.* Sci Rep, 2021. **11**(1): p. 22735.
72. Dupuis, N.F., E.D. Holmstrom, and D.J. Nesbitt, *Molecular-crowding effects on single-molecule RNA folding/unfolding thermodynamics and kinetics.* Proceedings of the National Academy of Sciences, 2014. **111**(23): p. 8464-8469.
73. Ji, N., et al., *Exploring the interaction of G-quadruplex and porphyrin derivative by single protein nanopore sensing interface.* Analytica Chimica Acta, 2020. **1106**: p. 126-132.
74. Morris, M.J., et al., *The porphyrin TmPyP4 unfolds the extremely stable G-quadruplex in MT3-MMP mRNA and alleviates its repressive effect to enhance translation in eukaryotic cells.* Nucleic Acids Res, 2012. **40**(9): p. 4137-45.
75. Zamiri, B., et al., *TMPyP4 Porphyrin Distorts RNA G-quadruplex Structures of the Disease-associated r(GGGGCC)n Repeat of the C9orf72 Gene and Blocks Interaction of RNA-binding Proteins *.* Journal of Biological Chemistry, 2014. **289**(8): p. 4653-4659.
76. Hernández-Díaz, T., F. Valiente-Echeverría, and R. Soto-Rifo, *RNA Helicase DDX3: A Double-Edged Sword for Viral Replication and Immune Signaling.* Microorganisms, 2021. **9**(6).
77. Xu, S., et al., *Thioflavin T as an efficient fluorescence sensor for selective recognition of RNA G-quadruplexes.* Sci Rep, 2016. **6**: p. 24793.
78. Herdy, B., et al., *Analysis of NRAS RNA G-quadruplex binding proteins reveals DDX3X as a novel interactor of cellular G-quadruplex containing transcripts.* Nucleic Acids Res, 2018. **46**(21): p. 11592-11604.

79. Caterino, M. and K. Paeschke, *Action and function of helicases on RNA G-quadruplexes.* Methods, 2022. **204**: p. 110-125.
80. Liu, H., et al., *A Helicase Unwinds Hexanucleotide Repeat RNA G-quadruplexes and Facilitates Repeat-Associated Non-AUG Translation.* Journal of the American Chemical Society, 2021. **143**(19): p. 7368-7379.
81. Bryan, T.M., *G-quadruplexes at Telomeres: Friend or Foe?* Molecules, 2020. **25**(16).
82. Moon, S.L., et al., *Flavivirus sfRNA suppresses antiviral RNA interference in cultured cells and mosquitoes and directly interacts with the RNAi machinery.* Virology, 2015. **485**: p. 322-9.
83. Lista, M.J., et al., *Nucleolin directly mediates Epstein-Barr virus immune evasion through binding to G-quadruplexes of EBNA1 mRNA.* Nat Commun, 2017. **8**: p. 16043.
84. Reznichenko, O., et al., *Novel cationic bis(acylhydrazones) as modulators of Epstein-Barr virus immune evasion acting through disruption of interaction between nucleolin and G-quadruplexes of EBNA1 mRNA.* Eur J Med Chem, 2019. **178**: p. 13-29.
85. Michalski, D., et al., *Zika virus noncoding sfRNAs sequester multiple host-derived RNA-binding proteins and modulate mRNA decay and splicing during infection.* Journal of Biological Chemistry, 2019. **294**(44): p. 16282-16296.
86. Fleming, A.M., N.L.B. Nguyen, and C.J. Burrows, *Colocalization of m(6)A and G-quadruplex-Forming Sequences in Viral RNA (HIV, Zika, Hepatitis B, and SV40) Suggests Topological Control of Adenosine N (6)-Methylation.* ACS Cent Sci, 2019. **5**(2): p. 218-228.

Preface: ***This chapter was published in the Journal Medical Virology in Wiley Online Library, Volume 95, Issue 5, May 2023, and is in its standard journalistic format.*** **Danielle Gemmill performed all G4 purification and folding preparation, CD spectroscopy, preparation for SAXS, Thioflavin T assay, DHX36 expression and purification, DHX36 and TMPyP4 binding affinity experiments with MST, and performed the DHX36 and TMPyP4 competing assays alongside Higor Sette Pereira. I additionally co-wrote the introduction to the manuscript, as well as the CD spectroscopy and MST sections. Higor Sette Pereira wrote the rest of the manuscript, as well as performed the SAXS data analysis and structure determination, and performed alongside M. Quadir Siddiqui, the mammalian cell culture experimentation with the MKPX A27L and A50R genes and RFP reporter system. Gunjan Vasudeva assisted in genomic analysis, as well as assisting in editing the manuscript, and Trushar R. Patel provided insight and oversaw the project.**

Chapter 7. Mapping and Characterization of G-quadruplexes in Monkeypox Genomes

Danielle Gemmill[1#], Higor Sette Pereira[1#], M. Quadir Siddiqui[1], Gunjan Vasudeva[1], Trushar R. Patel[1,2,3*]

[1] Alberta RNA Research and Training Institute & Department of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, AB T1K 3M4, Canada

[2] Department of Microbiology, Immunology and Infectious Disease, Cumming School of Medicine, University of Calgary, Calgary, AB T2N 4N1, Canada.

[3] Li Ka Shing Institute of Virology, University of Alberta, Edmonton, AB T6G 2E1, Canada.

[#] These authors contributed equally to this work

[*] **Corresponding author**: Trushar R. Patel (trushar.patel@uleth.ca)

7.1 Abstract

Monkeypox virus (MPXV) is a double-stranded DNA virus from the family *Poxviridae*, which is endemic in West and Central Africa. Various human outbreaks occurred in the 1980s, resulting from a cessation of smallpox vaccination. Recently, MPXV cases have re-emerged in non-endemic nations, and the 2022 outbreak has been declared a public health emergency. No treatment option is available, and many countries lack the infrastructure to provide symptomatic treatments. The development of cost-effective antivirals could ease severe health outcomes. G-quadruplexes have been a target of interest in treating viral infections with different chemicals. In the present work, a genomic-scale mapping of different MPXV isolates highlighted two conserved putative quadruplex-forming sequences MPXV-exclusive in 590 isolates. Subsequently, we assessed the G-quadruplex formation using CD spectroscopy and SAXS. Furthermore, biochemical assays indicated the ability of MPXV quadruplexes to be recognized by two specific G4-binding partners - Thioflavin T and DHX36. Additionally, our work also suggests that a quadruplex binding small molecule with previously reported antiviral activity, TMPyP4, interacts with MPXV G-quadruplexes with nanomolar affinity in the presence and absence of DHX36. Finally, cell biology experiments suggest that TMPyP4 treatment substantially reduced gene expression of MPXV proteins. In summary, our work provides insights into the G-quadruplexes from the MPXV genome that can be further exploited to develop therapeutics.

7.2 Introduction

Monkeypox virus (MPXV) belongs to the *Poxviridae* family and *Orthopoxvirus* genus, with the smallpox virus as a popular close-related human pathogen. Primarily transmitted through direct contact with infected humans or animal reservoirs, MPXV triggers a smallpox-like illness distinguished by premature enlargement of lymph nodes [1]. MPXV is a zoonotic vesicular-pustular malady first identified in humans in the Democratic Republic of Congo (DRC) in 1970 [2]. Since then, the infection has been considered endemic in DRC, propagating the cases to central and western African nations [3]. Initially thought to be a Southern Hemisphere geographical disease, MPXV's first infected case in the midwestern United States was detected in 2003 [4]. Although previously deemed a rare disease, MPXV has emerged as the most critical human *Orthopoxvirus* infection since the eradication of *Smallpox* in 1977 [1], and it is evolving into a global public health threat with increasing numbers of cases in more than 30 countries outside of Africa [5]. Currently, no approved treatment exists exclusively for MPXV and patients are guided to symptomatic medication [6]. Even though *Smallpox* vaccines can offer 85% protection against MPXV [7], a continuous vaccination program was discouraged after determining the virus was incapable of maintaining itself in human populations [8]. An absence of surveillance measures and cessation of smallpox vaccination have weakened herd immunity and are likely contributing to the re-emergence in non-endemic areas [3, 6]. In its latest report, the Centers for Disease Control (CDC) and Prevention confirmed a total of 86,231 Monkeypox cases, with 98% of the cases identified in previously unaffected regions and locations.

Poxviridae family members carry a sizeable double-strand DNA genome in approximately 200 kbp that multiply into the host cell and are organized within a Central conserved region flanked by left and right terminal regions that hold inverted terminal repeats (ITRs) [7]. MPXV's large genome equivalently contains multiple open reading frames that encode the entire proteins responsible for viral replication, transcription, and virion assembly [7]. Clade I and Clade II, formerly known as Congo Basin and West Africa respectively, compartmentalize MPXV genomes, and no meaningful reduction in size and content was observed by examining new draft genomes released from 2022 isolates [6]. However, structural features present in the MPXV genome are still poorly investigated, and most of the currently known details were deciphered in a comparison based on close-related *Poxviridae* representatives. Recent research shed light on an RNA quadruplex structure in the C9L gene that became unstable during the evolution course, which was believed to be a significant element in MPXV transmission [9].

G-quadruplexes (G4s) are non-canonical DNA and RNA structures formed by guanine-rich sequences that fold into two or more consecutive quartets [10]. A G-quartet refers to four planar guanines separated in two equivalent tracts collectively interacting via Watson-Crick and Hoogsteen hydrogen bonds. G4 is formed by stacking two or more quartets and preferentially stabilized by monovalent cations ($K^+ > Na^+ > Li^+$) and disturbed by the presence of divalent cations, like Mg^{2+} and Ca^{2+} [11, 12]. DNA quadruplexes can fold into parallel, antiparallel, and hybrid topologies based on strand direction, likewise loop size and composition [10]. To facilitate computational screening of putative quadruplexes, a theoretical consensus sequence was established as follows G_{3+}-N_{1-7}-G_{3+}-N_{1-7}-G_{3+}-N_{1-7}-G_{3+}-N_{1-7}, where G is the guanosine tracts, and N is the nucleotide content inside the loop [13]. However, quite a few DNA and RNA quadruplexes were discovered and characterized in literature uncompliant within this consensus sequence [14, 15].

These non-canonical G4 structures have been identified in all domains of life, modulating critical human processes, such as replication, transcription, mRNA splicing, translation, and epigenetic regulation of the genome, as previously reviewed [16, 17]. G4s have also been discovered across microbial genomes that play an essential role in facilitating immune evasion, radioresistance, virion secretion, and nitrate metabolism [18]. Targeting G4s is a promising strategy to control microbial infection by inhibiting critical pathways. By specifically targeting RNA G4, researchers identified a reduced level of Ebola virus L gene expression in a mini-genome system that impacts decreasing viral replication [19]. Moreover, in Herpes Simplex Virus-1 infected cells, DNA G4s were shown to participate in antiviral cellular defence pathways, and once those structures are stabilized, viral infection is diminished to controlled levels [20]. RNA G4s were also targeted in Zika virus-infected cells and shown to mitigate viral levels acting as a roadblock stalling polymerase and ribosome processing [21]. Interestingly, the abovementioned papers used a cationic porphyrin, TMPyP4, a potential inhibitor of human telomerase [22], to target different G4s structures, indicating its antiviral activity.

In this work, a robust genomic screening in MPXV isolates collected during different outbreaks was performed to map G-rich sequences. Subsequently, we utilized Quadruplex-forming G-rich Sequences (QGRS) Mapper [23] and G4RNA screener [24] to calculate putative G4 folding scores of G-rich sequences from the MPXV genome. A conservation study indicated two unique putative G4 structures disturbed in close-related *Orthopoxvirus*. Using biophysical tools, we further characterized both G-rich sequences to establish that they adopt G4 structures in solution. Finally, we demonstrate the ability of G4s to interact with TMPyP4 *in vitro*. Overall, our results demonstrate for the first time that the MPXV genome contains G4-forming sequences that could potentially be targeted using TMPyP4 and other G4-interacting drugs to explore therapeutic development avenues.

7.3 Materials and Methods

7.3.1 Genome mapping and sequence conservation analysis

Genome data were collected on the Nucleotide database of NCBI, where all 590 complete MPXV genomes (accessed in August 2022) isolated from different countries and years were assembled for further sequence analysis. To assess the diversity of sequences, a phylogenetic tree using the 590 sequences was performed as following. MPXV genome file was aligned using MAFFT (Multiple Alignment using Fast Fourier Transform)[25] and further assigned to different clades according to differences in query and reference sequences through Nextclade. The maximum-likelihood method was used to construct a phylogenetic tree using IQ-tree [26] and the tree was finally visualized using iTOl v5[27]. A total of 38 sequences were assigned to Clade I, whereas 552 sequences are accommodated in Clade II. Then, the Putative Hybrid-Quadruplex forming Sequence (PHQS) script performed an initial screening to map G-rich regions in all genomes as previously described [28] according to the following parameters: G-tract size = 2, loop size 1 to 5, number of g-tracts equal or higher than 4, search range 210,000, bin size 100 nucleotides. The percentage of hits was achieved by dividing the number of times each sequence appeared in the prediction for the number of genomes used (n=590). Next, recovered sequences were ranked using QGRS mapper [23], following default parameters, and G4RNA screener [24] based on cGcG (Consecutive G over consecutive C ratio), G4H (G4 Hunter), and G4NN (G4 Neural Network) with a defined threshold of cGcG >4.5, G4H >0.9, and G4NN >0.5.

The nucleotide Basic Local Alignment Tool (BLASTn) [29] was used to assess similarities within MPXV sequences against viruses from the *Orthopoxvirus* genus. MPXV selected putative G4-sequences were individually aligned with *Smallpox* (taxid:10255, n=55), *Cowpox* (taxid:10243, n=22), *Camelpox* (taxid:28873, n=9), and *Horsepox* (taxid:397342, n=3) genomes using BLASTn default parameters. Matching sequences were further re-aligned using Clustal Omega multiple sequence alignment tool [30].

Table 8.1 Representative codes and sequences for the oligos used in this study.

Oligo	Sequence
MP1	5’ ATTA*GG*T*GG*G*GG*AT*GG*ACAA 3’
MP1 mutant	5’ ATTA*GG*TGAGAGAT*GG*ACAA 3’
MP2	5’ GGTAAA*GG*A*GG*AAA*GG*GT*GG* 3’
MP2 mutant	5’ GATAAA*GG*AGAAAAA*GG*T*GG* 3’
T7 primer	5’ TAATACGACTCACTATAGGG 3’
Telomeric G4	5’ TTA*GGG*TTA*GGG*TTA*GGG*TTA*GGG* 3’

7.3.2 Circular dichroism (CD) spectroscopy

Synthetic lyophilized oligonucleotides were purchased from Alpha DNA (Montreal, Quebec). The oligos used in this study are described in **Table 7.1**. MP1, MP2 and their respective mutants were dissolved to 20 µM using a G4 buffer (20 mM HEPES, 100 mM KCl, and 0.2 mM EDTA, at pH 7.4). Before data collection, the oligos were heated at 95 °C for 5 minutes and cooled down at room temperature for 30 minutes. Jasco J-815 spectropolarimeter (Jasco Inc, Easton, MD) was used to collect spectra of all oligos ranging from 220 to 320 nm, using a 1.0 mm cell, 0.1 nm data pitch, with five accumulations and 32 s integration time. A continuous supply of nitrogen gas was provided to avoid water condensation in the cells. All measurements were baseline corrected with G4 buffer and repeated in triplicate.

7.3.3 Solution small-angle X-ray scattering (SAXS)

SAXS data were collected using the B21 BioSAXS beamline at the Diamond Light Source (Oxfordshire, UK), using an Agilent 1200 HPLC connected in line with a specialized flow-cell, as previously described [31]. Samples of wild-type and mutant oligos were concentrated to 150 µM, heated and cooled as described above, and injected into the Shodex KW402.5-4F, pre-equilibrated with G4 buffer. About 600 frames, with 3seconds exposure were collected.

Scattering data were analyzed using the ATSAS suite, version 2.8 [32]. We used Chromixs to buffer the baseline correct samples peak intensity [33]. Samples' quality check and the radius of gyration (R_g) were assessed by performing a Guinier analysis (q^2 vs. $\ln(I(q))$) [34]. The folding of wild-type and mutants were obtained by a dimensionless Kratky plot (qR_g vs $qR_g^2 * I(q)/I(0)$) [35] and a paired-distribution function (P(r)) provided the real-space R_g and the maximum particle dimension (D_{max}) using GNOM [36]. P(r) plot was the input data to DAMMIN [37] calculates eleven models of each oligo, selecting different seeds for each model although using equal parameters to ensure symmetry. DAMAVER was used to obtain an average model, and DAMFILT yielded a filtered representative structure model [38].

7.3.4 Microscale thermophoresis (MST) studies

MST experiments were performed at room temperature using the Nanotemper Technologies Monolith® NT.115 (Munich, Germany) instrument and standard capillaries to measure fluorescence counts, binding affinity, and binding check. For the Thioflavin assay, 5 µM of oligos diluted in G4 buffer and previously heat-cooled (as mentioned above) were incubated with 1 µM of Thioflavin T for 15 minutes. MST was used to collect initial fluorescence counts. Samples were analyzed in quadruplicate; fluorescence counts were normalized based on telomeric G4 (positive control, sequence in **Table 7.1**), and data were statistically analyzed using an unpaired t-test on GraphPad Prism 9.4 software.

A DHX36 construct carrying the N-terminal amino acid residues ranging from 53-105 (DHX36$_{53\text{-}105}$, containing DNA and RNA G4 interacting motif, was cloned and purified as previously described by our group [14]. The reaction was prepared by mixing 100 nM of 5' FITC labelled oligonucleotide with DHX36$_{53\text{-}105}$ ranging from 0.001 to 17.5 µM diluted in MST buffer (20 mM HEPES, 100 mM KCl, 0.2 mM EDTA, and 0.05% Tween 20, at pH 7.4), and incubated at room temperature for 20 minutes. Subsequently, binding affinity experiments were performed using medium MST-Power, 50% excitation-power, with data collection on the cold region at 0 seconds and the hot region at 5 seconds.

Another binding affinity experiment tested the interaction between TMPyP4 within G4-forming and mutant oligos following the same methodology described above. TMPyP4 was acquired from Sigma Aldrich (CAS number 36951-72-1, Massachusetts, USA), diluted from 1.1 to 0.0003 µM and further incubated at room temperature for 20 minutes with 100 nM of 5' FITC labelled oligonucleotides. Three independent replicates from DHX36$_{53\text{-}105}$ and TMPyP4 binding affinity were submitted to data collection and analyzed on MO Affinity Analysis software (version 2.1.3). The software uses thermophoresis-based fluorescence changes to calculate the binding affinity, K_d. GraphPad Prism 9.4 software was used to plot curves.

For the competitive assay, 100 nM of G4-forming oligonucleotides were mixed first with 500 nM of TMPyP4 in MST buffer; the reaction was prepared in four tubes and incubated at room temperature for 20 minutes. Then, 500 nM of DHX36$_{53\text{-}105}$ was added to two tubes, whereas the other two were filled with MST buffer. In a different set, the oligonucleotides were first incubated with 500 nM of DHX36$_{53\text{-}105}$ for 30 minutes, and then half of them were incubated with 500 nM of TMPyP4 for an additional 30 minutes. The experiment was performed in duplicate with four reactions in each set, totaling eight points per condition tested. Data was collected via an MST binding check that ran using medium MST power, 50% excitation power, on the cold region at 0 seconds and the hot region at 5 seconds at room temperature. Fluorescence changes due to thermophoresis were recorded and normalized using MO Affinity Analysis software (version 2.1.3), and the graph was plotted using GraphPad Prism 9.4 software.

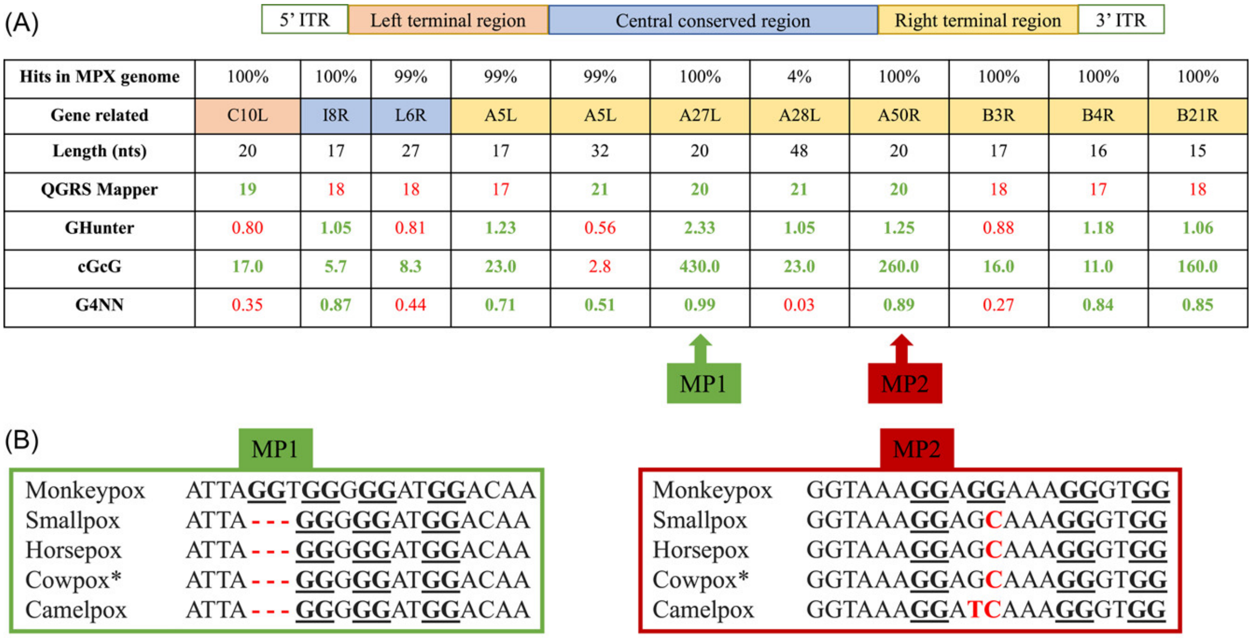
7.3.5 MPXV gene expression experiments

The MPXV genes A27L and A50R were cloned into a mammalian expression vector, pcDNA3.1(+), and expressed linked with red-fluorescence protein (RFP) as a reporter. The A27L and A50R sequences were retrieved from the *Monkeypox* reference sequence (NC_063383) available in the NCBI database and synthesized and purchased from Biomatik (Ontario, Canada). An RFP-expressing plasmid (#26720, Addgene) was used as a negative control. The plasmids were transformed into *Escherichia coli* DH5α and selected on ampicillin-resistant agar plates. Individual colonies were inoculated in LB media and plasmids were purified using the GeneJET Plasmid Maxiprep Kit (ThermoFisher, Canada) according to the manufacturer's instructions. HEK293 cells were seeded in 35 mm petri dishes (2.0 x 10^5 cells/dish). Once the cells reached ~80% confluence, they were transfected with 5 µg of A27L, A50R, or RFP_{ctrl} plasmids per dish using Lipofectamine™ 3000 Transfection Reagent (Invitrogen, Canada) following the manufacturer's protocol. After 24 hours of transfection, cells were treated with 0, 10, 15, or 20 µM of TMPyP4 and kept in the dark for an additional 20 hours. The cells were then imaged using an Olympus FV1200 confocal microscope with brightfield microscopy and RFP filter, using excitation 488 nm and emission 561 nm. Images were collected and analyzed using FV1200 Olympus software.

7.4 Results

7.4.1 Predicted G4 sequences are predominantly localized in the Right terminal region and are MPXV exclusive.

We employed a computational screening in 590 genome isolates across the globe using the PHQS script exploring G-rich sequences following a pattern of $(G\geq 2, 1\leq N\geq 5)_{4\leq}$, where G is guanines, and N is the number of nucleotides held in the loop between two G-tracts. Eleven putative G-rich sequences were retrieved, with seven identified in all 590 genomes (100% hits), three in 588 genomes (99%), and one only detected in 24 genomes (4%) (**Figure 7.1A**). MPXV genome carries left and right terminal regions that comprise non-coding Inverted Terminal repeats (ITR). One putative sequence was predicted on the Left terminal region, whereas eight sequences are concentrated at the Right terminal region. Also, a Central conserved region contains two putative sequences in the middle of the genome. Using the Democratic Republic of Congo 2008 isolate as reference (GenBank accession KP849469), a corresponding gene for each putative sequence was identified. All predicted sequences are localized in the coding region, and their respective protein code is available in **Figure 7.1A**. The nucleotide length varies from 15 to 48, but the G-score was calculated using the minimal predicted G4. G-scores above the threshold are presented in bold green, whereas below the threshold are presented in red. QGRS Mapper identified four putative G4-sequences (threshold ≥ 19) while Ghunter (threshold ≥ 0.9), cGcG (threshold ≥ 4.5) and G4NN (threshold ≥ 0.5) detected seven, ten, and seven sequences, respectively. Collectively analyzing the predicted G-score along with conservation hits across MPXV strains, G4-sequences related to the genes A27L and A50R were selected for further studies and referred to as MP1 and MP2, respectively.



Hits in MPX genome	100%	100%	99%	99%	99%	100%	4%	100%	100%	100%	100%
Gene related	C10L	I8R	L6R	A5L	A5L	A27L	A28L	A50R	B3R	B4R	B21R
Length (nts)	20	17	27	17	32	20	48	20	17	16	15
QGRS Mapper	19	18	18	17	21	20	21	20	18	17	18
GHunter	0.80	1.05	0.81	1.23	0.56	2.33	1.05	1.25	0.88	1.18	1.06
cGcG	17.0	5.7	8.3	23.0	2.8	430.0	23.0	260.0	16.0	11.0	160.0
G4NN	0.35	0.87	0.44	0.71	0.51	0.99	0.03	0.89	0.27	0.84	0.85

MP1

Monkeypox	ATTA**GG**T**GG**G**GG**AT**GG**ACAA
Smallpox	ATTA- - - **GG**G**GG**AT**GG**ACAA
Horsepox	ATTA- - - **GG**G**GG**AT**GG**ACAA
Cowpox*	ATTA- - - **GG**G**GG**AT**GG**ACAA
Camelpox	ATTA- - - **GG**G**GG**AT**GG**ACAA

MP2

Monkeypox	GGTAAA**GG**A**GG**AAA**GG**GT**GG**
Smallpox	GGTAAA**GG**AGCAAA**GG**GT**GG**
Horsepox	GGTAAA**GG**AGCAAA**GG**GT**GG**
Cowpox*	GGTAAA**GG**AGCAAA**GG**GT**GG**
Camelpox	GGTAAA**GG**ATCAAA**GG**GT**GG**

Figure 7.1 Mapping putative quadruplexes in MPXV genome followed by a conservation analysis. A) Representative genome of the Poxviridae family flanked by 5' and 3' Inverted Terminal repeats (ITR) followed by a Left terminal region (orange), Central conserved region (blue) and Right terminal region (yellow). Gene localization was individually represented in each column following the color code used before to demonstrate genomic position. Then, the nucleotide length of putative sequences and predicted G-score for each are presented in the subsequent rows. Numbers displayed in bold green are above the defined threshold, whereas red numbers correspond to G-scores under the threshold. B) Conservation analysis of MP1 (green square) and MP2 (red square) sequences against human disease related Orthopoxvirus. *Only Cowpox has one genome with 100% similarity with MP1 and four with 100% similarity with MP2.

A local alignment was conducted to analyze the conservation of selected putative G4-sequences against disease-associated members of the *Orthopoxvirus* genus (**Figure 7.1B**). MP1 and MP2 were aligned with *Smallpox* (n=55), *Cowpox* (n=22), *Camelpox* (n=9), and *Horsepox* (n=3), and their corresponding sequences are presented in **Figure 7.1B**. Sequence alignment for MP1 demonstrates 85% similarity, with three nucleotide gaps in the center. Regarding MP2, *Smallpox*, *Horsepox* and *Cowpox* contain a cytosine residue instead of a guanine residue in the

MPXV genome, whereas *Camelpox* differs from MPXV by substituting double guanine residues with thymidine and cytosine residues. Cowpox has one genome with 100% similarity with MP1 (DQ437593.1) and four with 100% similarity with MP2 (KY549143.2, KY369926.1, HQ407377.1, X94355.2). Interestingly all variations were observed to be substituting guanine residue disrupting G-tracts, negatively impacting the putative G4-sequence score.

7.4.2 Biophysical evaluation of G4 formation

To assess the ability of MP1 and MP2 to fold into a G4, two 20-mer wild-types (MP1 and MP2) oligos were synthesized and evaluated using biophysical methods (**Table 7.1**). Additionally, MP1 and MP2 mutants were also designed to disrupt the G4 second quartet by replacing two guanines with adenines. Immediately before performing experiments, all the oligonucleotides were heated at 95 °C for 5 minutes and cooled down at room temperature for at least 15 minutes. CD spectroscopy was used to investigate G4 folding and topology in the presence of potassium ions (**Figure 7.2A**). A noticeable positive peak at around 263 nm and a deep negative peak at about 243 nm were observed for MP1 (left plot, continuous green line) and MP2 (right plot, continuous red line), suggesting that both oligonucleotides can fold into a G4 structure in a buffer containing potassium ions and adopt a parallel topology [39]. Inversely, the MP1 mutant (left plot, dashed green line) showed a shift to the right with a positive peak around 272 nm and a negative close to 245 nm. MP2 also had a minor right change in the negative peak to around 248 nm, and a broad positive peak could be seen from 260 until 275 nm. Both mutant patterns are atypical to any well-known G4 topology, suggesting that these oligos cannot fold into a G4 structure.

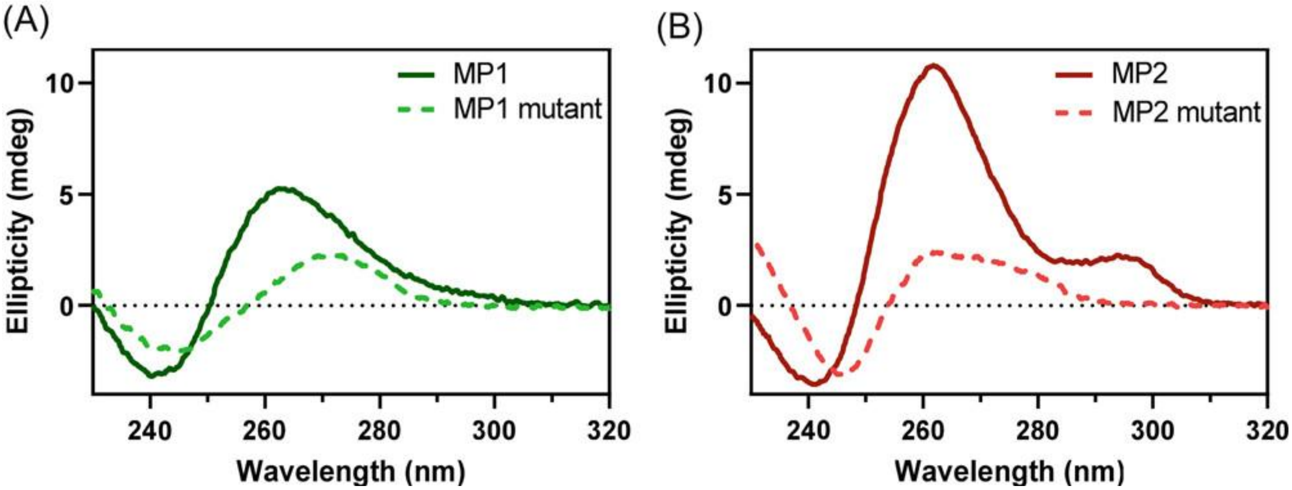


Figure 7.2 Circular dichroism spectroscopy reveals that MPXV oligos fold in a parallel topology. A) and B) circular dichroism spectroscopy profiles of MPXV G4 MP1 (continuous green line) and MP2 (continuous red line), as well as MP1 mutant (dashed green line) and MP2 mutant (dashed red line), demonstrate that a single mutation in MP1 and MP2 G4 oligos leads to change in G4 secondary structures.

Further, low-resolution structures were determined to visualize differences in wild-type G4-forming oligos compared to their respective mutants using SAXS. The merged intensity data and Guinier analysis (**Figure 7.3A**) demonstrate that our samples are pure and monodisperse. The linearity observed in the Guinier plots indicates an aggregation-free solution and allows the calculation of R_g using low angle region, presented in **Table 7.2** as Guinier R_g (Å). An elongated structure could be observed in both mutants with R_g of approximately 19 Å, whereas MP1 and MP2 are more compacted, with 17.18 and 16.31 Å, respectively. Next, we performed dimensionless Kratky and P(r) function analysis to further gain insights into oligos' folding, conformation, and compactness states [40]. The dimensionless Kratky plot suggested that wild-type oligos exhibit comparable folding states, reasonably distinct from those steady plateau displayed by mutants but indicating that all oligos are folded and are appropriate for low-resolution three-dimensional structure determination (**Figure 7.3B**).

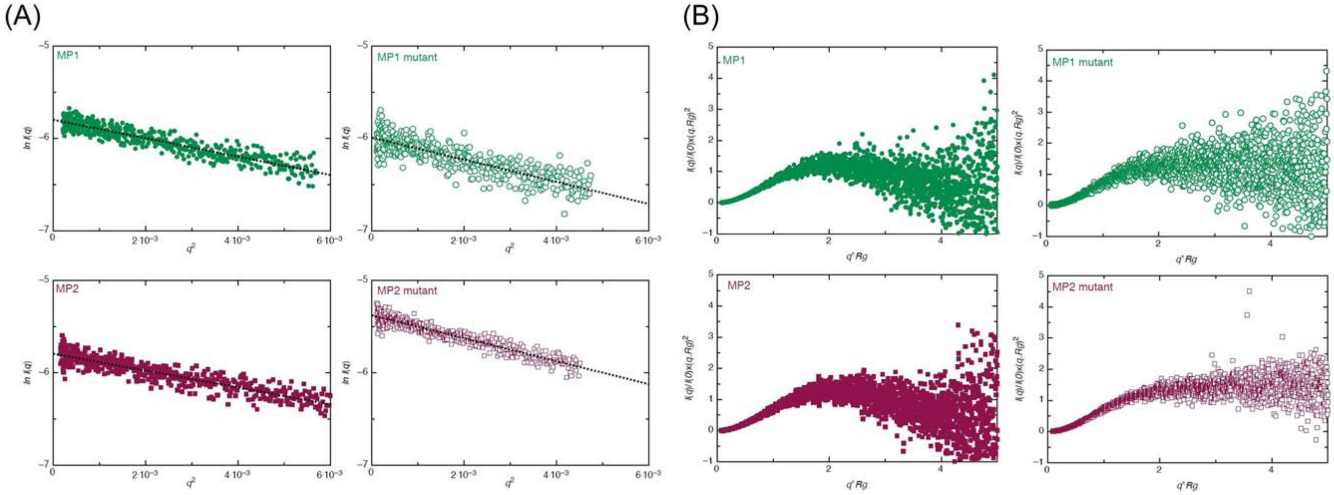


Figure 7.3 Guinier and Kratky's plots indicate that oligomers are free of aggregation and properly folded in solution. A) A liner Guinier (Ln(I,q)) vs q^2) region indicates that all samples are pure and aggregation-free. B) Dimensionless Kratky plots (I(q)/I(0)*(q*R_g)^2 vs. q*R_g) suggest that both wild-type and mutants are folded and appropriate for low-resolution structure determination. Also, an effect of a single mutation for MP1 and MP2 G4 oligo is evident by the differences between dimensionless Kratky plots for wild-type and mutants. MP1 = green colored circle, MP1 mutant = green blank circle, MP2 = red colored square, and MP2 mutant = red blank square.

Table 7.2 Analysis of small-angle X-ray scattering data for MP1, MP2, and their mutants.

Parameters	**MP1**	**MP1 mutant**	**MP2**	**MP2 mutant**
Sequence Molecular Weight (gmol^{-1})	6,286.2	6,254.2	6,360.2	6,312.2
Guinier I(0)	0.0031 ± 1.7x10^{-5}	0.0025 ± 2x10^{-5}	0.0031 ± 1.6x10^{-5}	0.0047 ± 2x10^{-5}
Guinier R_g (Å)	17.18 ± 0.20	18.90 ± 0.30	16.31 ± 0.17	19.38 ± 0.16
q.R_g range	0.25 – 1.30	0.20 – 1.30	0.20 – 1.30	0.24 – 1.30
P(r) I(0)	0.0030 ± 1.1x10^{-5}	0.0024 ± 1.5x10^{-5}	0.0030 ± 1.6x10^{-5}	0.0046 ± 1.6x10^{-5}
P(r) R_g(Å)	16.18 ± 0.05	18. 28 ± 0.10	15.54 ± 0.05	19.55 ± 0.10
D_{max} (Å)	44	55	45	59
χ^2	1.10	1.05	1.18	1.08
NSD	0.64 ± 0.02	0.53 ± 0.01	0.66 ± 0.02	0.57 ± 0.01

The P(r) plot was derived from indirect Fourier transformations to convert reciprocal information on the Guinier plot into a real-space electron paired distribution [41]. Oligos' R_g was calculated in real-space (P(r) R_g(Å), in **Table 7.2**) and corresponded to those achieved by the Guinier plot, with slight differences in G4-forming oligos where real-space R_g are 16.18 and 15.54 Å, presented with about one Å unit more in Guinier. Additionally, Gaussian distribution observed in the P(r) plot indicates that both mutants have a compacted structure with equivalent maximum particle dimension, yet the mutants are displayed in an elongated structure with D_{max} around 55 and 59 Å for MP1 and MP2, respectively (**Figure 7.4**).

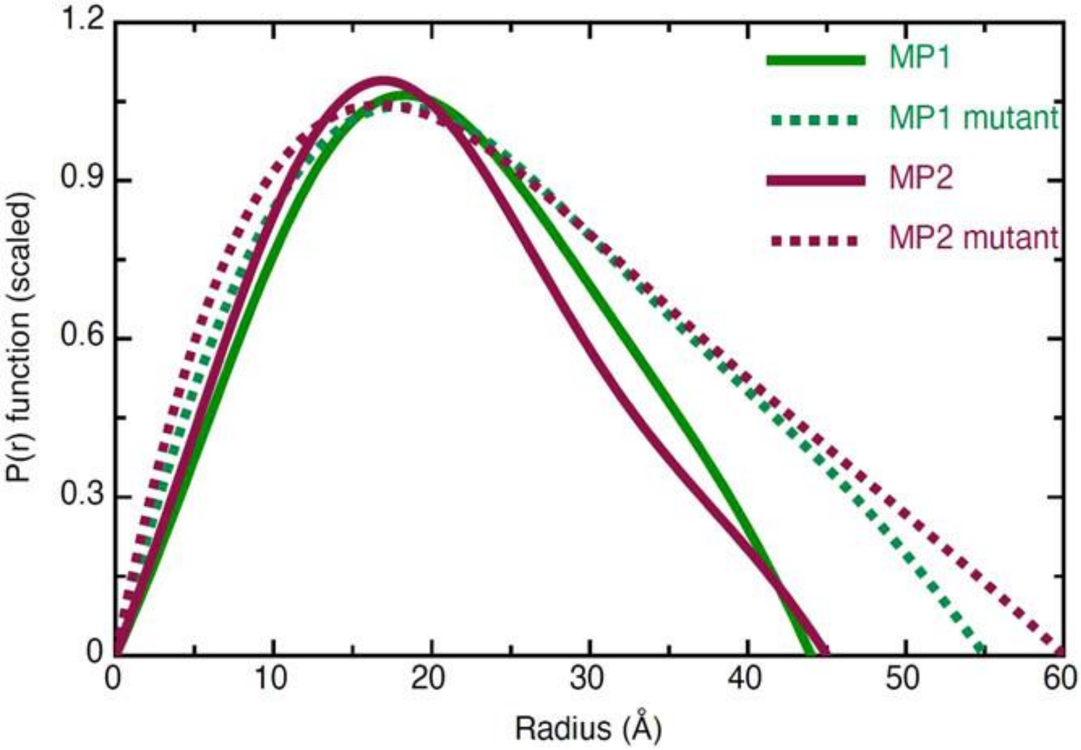


Figure 7.4 Normalized pair distribution function of MPXV oligomers. P(r) vs radial distance plot indicates that as expected, both wild-type G4s have similar dimensions and are compact, whereas, both mutants are extended in solution. MP1 (continuous red line) and MP2 (continuous green line) assume a compact and solid structure, while MP1 mutant (red dashed line) and MP2 mutant (green dashed line) fold into an elongated structure.

DAMMIN was used to generate eleven low-resolution structures of wild-type and mutant oligos. The goodness of fitting (χ^2) within all generated models is presented in **Table 7.2** and statistically indicates a favourable agreement for all four oligomers. Consecutively, the normalized spatial discrepancy (NSD) was calculated to verify the agreement within low-resolution structures calculated using DAMMIN. NSD values, presented in **Table 7.2,** indicate that in each case, all DAMMIN-derived structures are in good agreement with each other. Consequently, we present averaged and filtered low-resolution structures for each oligos in **Figure 7.5**. Altogether, the low-resolution structures emphasize that mutants adopt comparable elongated structures while G4-forming sequences fold into dense and quasi-globular structures.

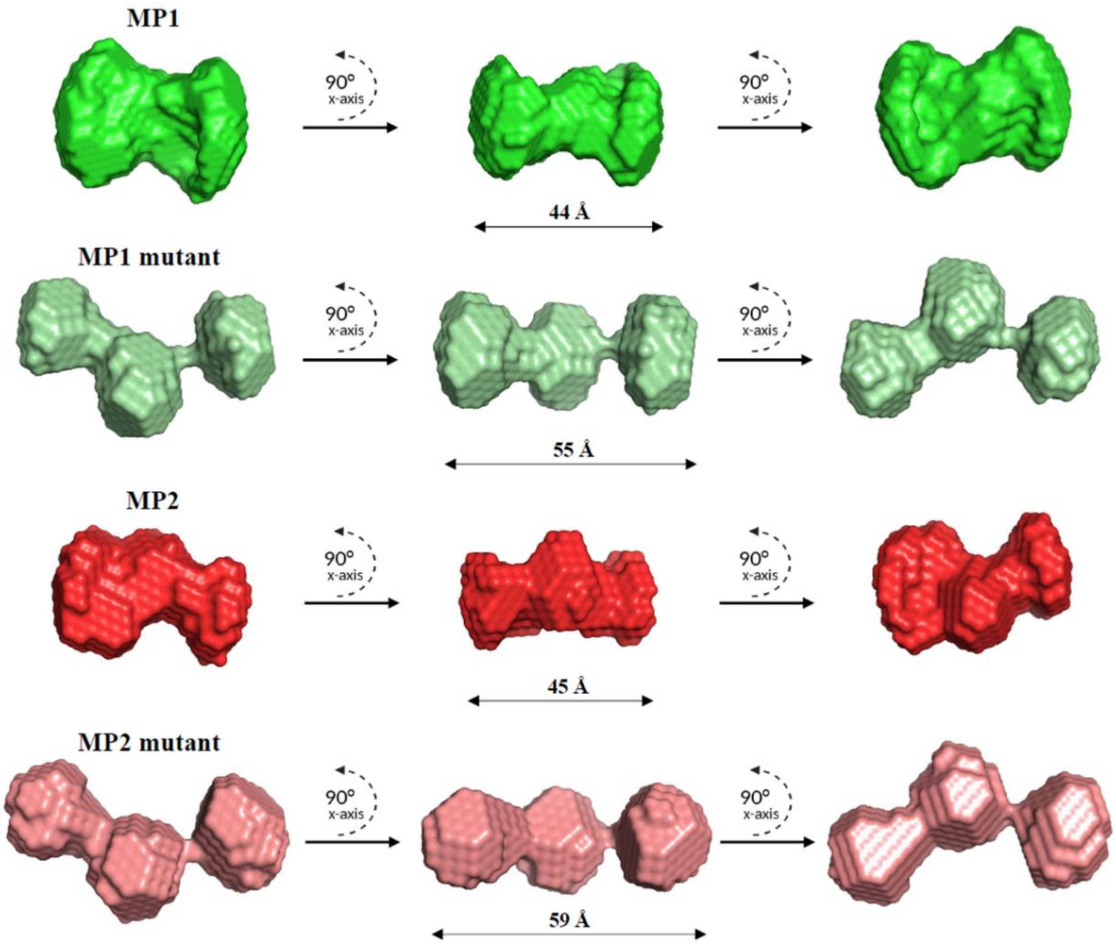


Figure 7.5 Low-resolution three-dimensional structures of MPXV oligos. DAMFILT filtered representative models with two consecutive 90° rotations along their x-axis. Horizontal arrows represent their D_{max} obtained from the pair-distribution function. As expected, MP1 (red) and MP2 (green) have a compacted quasi-globular shape, whereas MP1 mutant (pale green) and MP2 mutant (dark salmon) displayed an elongated structure that is very similar to each other.

7.4.3 Parallel G4s interact with quadruplexes-binding partners.

Next, we determined the ability of the oligomers to fold in solution and be recognized by well-established G4-binding partners. As previously demonstrated, Thioflavin T (ThT) recognizes telomeric G4 DNA with high affinity [42]. Therefore, relative fluorescence was calculated using Telomeric G4 as 100% fluorescence counts; T7 primer was used as a non-G4 forming sequence.

MP1 has an average relative fluorescence of 70% when compared to Telomeric G4, while MP1 mutant had a signal of around 35% (**Figure 7.6**). Also, it is perceptible that MP2 had a fluorescence signal of around 80% and its relative mutant a slightly superior intensity than T7 primer, at around 10%. An unpaired *t*-test statistically supports a significant distinction between MP1 and its mutant with a *p*-value = 0.001 and between MP2 and MP2 mutant, with the *p*-value < 0.001. Altogether, these results indicate that ThT is a suitable sensor for detecting MP1 and MP2 in solution, emphasizing their ability to fold into G4.

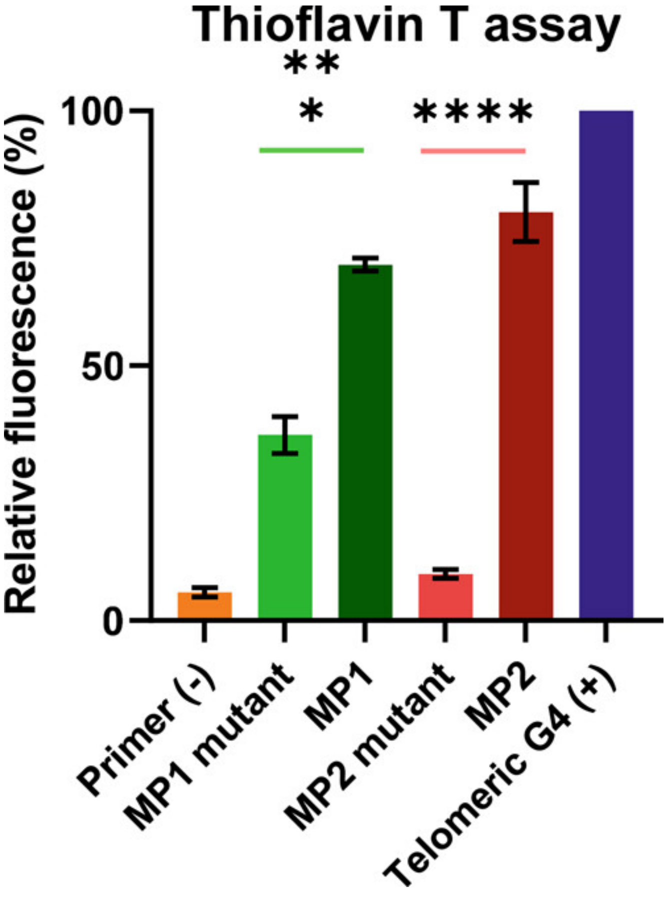


Figure 7.6 Thioflavin T assay indicates MPXV G4s folding by stacking two tetrads. Thioflavin T assay measuring the relative fluorescence of MPXV oligos, demonstrating that MP1 and MP2 present high amount of relative fluorescence, compared to non-G4 forming primers (negative control) or mutants. Relative values were calculated based on well-established Telomeric G4 (blue bar) as a positive control.

Furthermore, we also asked if DHX36$_{53\text{-}105}$, a member of the DEAH-box helicase family described to interact with G4 structures, can recognize MPXV G4s [14, 43, 44]. The MST binding assays of MPXV G4s with the G4 interacting N-terminal motif of DHX36$_{53\text{-}105}$ suggested the K_d of 44.0 ± 1.8 nM for MP1 (continuous green line) and 611.9 ± 234.3 nM for MP2 (continuous red line). Conversely, the binding affinities of mutants for the DHX36$_{53\text{-}105}$ were determined to be in a micromolar range (**Figure 7.7A**). Mutants from MP1 (green dashed line) and MP2 (red dashed line) interact with DHX36$_{53\text{-}105}$ with a binding affinity of 16,354.6 ± 1,632.7 nM and 11,953 ± 920.7 nM, respectively.

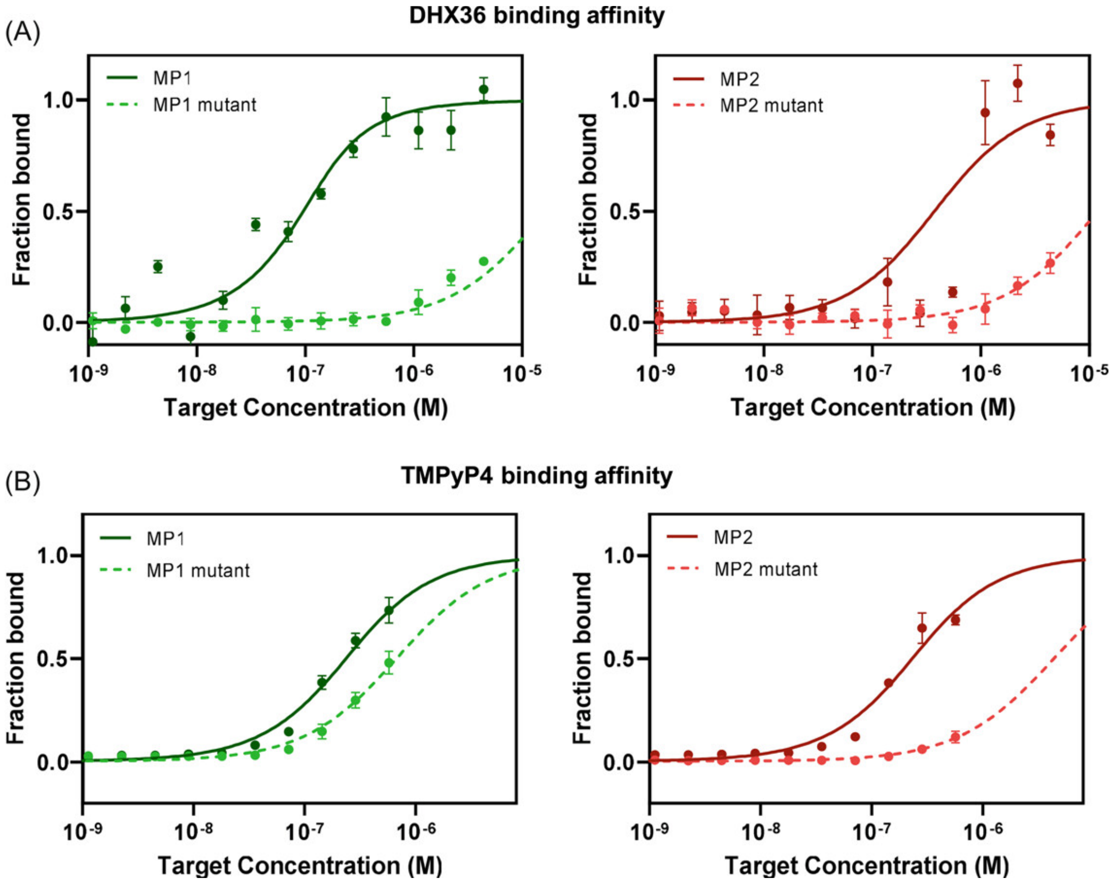


Figure 7.7 Assessment of binding affinity using MST between MPXV G4s and G4-binding partners. We determined the affinity of A) DHX36$_{53\text{-}105}$ and B) TMPyP4 with MPXV oligos. MST data for MP1 (continuous green line) and MP1 mutant (dashed green line) are presented on the

left side, and for MP2 (continuous red line) and MP2 mutant (dashed red line) on the right side. FITC fluorescent oligos were constantly added to the reaction capillaries while a varied concentration of DHX36$_{53\text{-}105}$ and TMPyP4 was tested. Dissociation constants (K_d) were calculated based on a curve fitting and are presented in **Table 7.2**. The binding curves demonstrate that both DHX36$_{53\text{-}105}$ and TMPyP4 interact with nanomolar affinity to MP1 and MP2, compared to mutants.

After utilizing a G4 interacting fluorescence sensor and DHX36$_{53\text{-}105}$ protein, we assessed if TMPyP4, a cationic porphyrin chemical that targets G4 structures, also recognizes MPXV G4s [20, 45]. We used MST to determine that MP1 and MP2 bind to TMPyP4 with K_d of 182.3 ± 14.3 nM and 180.9 ± 22.6 nM, respectively (**Figure 7.7B, Table 7.3**). MP1 mutant also interacts with TMPyP4 in a nanomolar range (612.3 ± 47.6 nM), although interaction strength is around 3.3 times weaker than MP1. Furthermore, the MP2 mutant interacts with TMPyP4 in a micromolar range (4,324.7 ± 199.2 nM), suggesting a binding event only starts at exceptionally high concentrations.

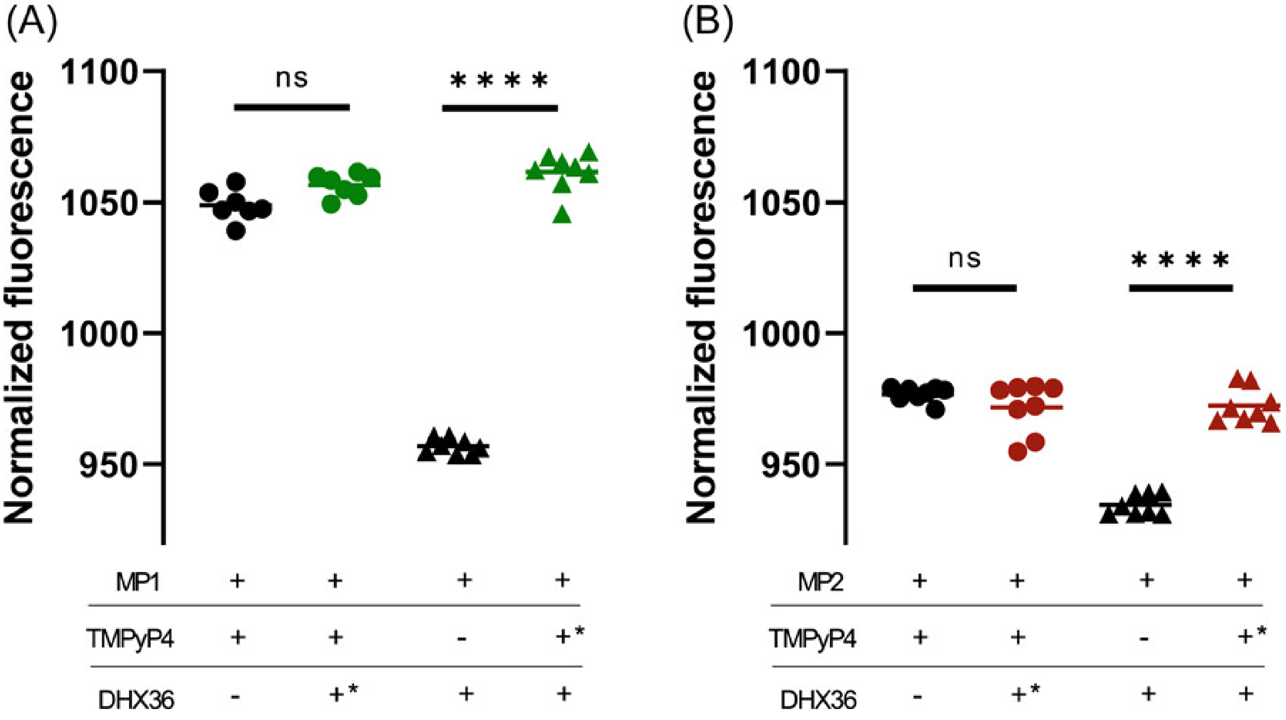


Figure 7.8 G4 binding partners competitive assay within MPXV G4s assessed using MST. A) MP1 and B) MP2 were first incubated with TMPyP4 in the absence of DHX36$_{53\text{-}105}$(black circles). In a different reaction, an equal concentration of DHX36$_{53\text{-}105}$ was added to the oligos and TMPyP4 mixture (MP1 – green circles in A) and MP2 – red circles in B)). Similarly, both oligos were incubated in the absence of TMPyP4 (black triangles) followed by an equivalent reaction with equal concentrations of TMPyP4 (MP1 – green triangles in A) and MP2 – red triangles in B)). These experiments demonstrate that MP1 and can be recognized by TMPyP4, even though they are bound to DHX36$_{53\text{-}105}$. On the other hand, if MP1 and MP2 is already bound to TMPyP4, they cannot interact with DHX36$_{53\text{-}105}$. Thus, these data highlight that TMPyP4 is able to interact with and stabilize MPXV G4s in the presence of host G4-interacting proteins. * Indicates the component that was secondly added to the reaction mixture.

Table 7.3 The dissociation constant (K_d) calculated using MST for MPXV G4's and their respective mutants' interaction with DHX3653−105 and TMPyP4.

	DHX36	**TMPyP4**
MP1	44.0 ± 1.8 nM	182.3 ± 14.3 nM
MP1 mutant	16,354.6 ± 1,632.7 nM	612.3 ± 47.6 nM
MP2	611.9 ± 234.3 nM	180.9 ± 22.6 nM
MP2 mutant	11,953.0 ± 920.7 nM	4,324.7 ± 199.2 nM

7.4.4 TMPyP4 outcompetes with DHX36$_{53\text{-}105}$ for MPXV G4 binding site

A competitive assay was performed to understand if TMPyP4 can recognize G4 structures in the presence of another intrinsic cellular partner, such as DHX36$_{53\text{-}105}$. Using MST, G4-forming oligos were incubated with TMPyP4 (**Figure 7.8**, black circles), followed by the determination of the normalized fluorescence for MP1 and MP2. Our assays indicated the fluorescence units of 1050 for MP1 and 975 for MP2 (**Figures 7.8A and 7.8B**). After adding an equal concentration of DHX36$_{53\text{-}105}$ to each reaction tube, no statistically significant differences were observed for both MP1 and MP2 (green and red circles in **Figure 7.8**). Inversely, G4-forming oligos were first incubated with DHX36$_{53\text{-}105}$ (black triangles in **Figure 7.8**) and then supplemented with an equal amount of TMPyP4. A significant statistical difference ($p < 0.0001$) was observed for MP1 (green triangles, **Figure 7.8A**) and MP2 (green triangles, **Figure 7.8B**) using an unpaired *t*-test. This competitive assay and interaction experiment strongly suggest that TMPyP4 outcompetes and stabilizes G4 structures, whether binding to an intrinsic cellular partner or not.

7.4.5 TMPyP4 treatment downregulates expression of MPXV A27L and A50R genes.

In order to determine whether TMPyP4 stabilization of G4s can modulate MPXV genes, we designed a gene expression system for MPXV A27L and A50R genes, which carry MP1 and MP2, respectively. MPXV gene expression was tracked through the fluorescence reporter RFP in a cellular system. A plasmid expressing RFP (RFP_{ctrl}) was used as a negative control upon TMPyP4 treatment, as no G4 has been reported on the RFP sequence. As expected, RFP expression in RFP_{ctrl} remained unaltered after treatment with TMPyP4 at different concentrations 10, 15 and 20 µM (**Figure 7.9**, top panel). However, treatment of MPXV genes caused a substantial reduction of RFP expression starting at 15 µM of the TMPyP4. While A27L gene expression (**Figure 7.9**, middle panel) was unaffected upon treatment with 10 µM of TMPyP4, however, a substantial reduction was observed at higher concentrations. A50R-RFP expression was observed through tiny fluorescence dots (**Figure 7.9**, bottom panel), in a different pattern than was noticed in RFP_{ctrl} and A27L. However, TMPyP4 treatment was effective in constraining A50R gene expression at similar concentrations to those observed in A27L. Additionally, bright microscope images demonstrated that cells are viable and healthy after treatment, suggesting non-toxic activity at the tested concentrations.

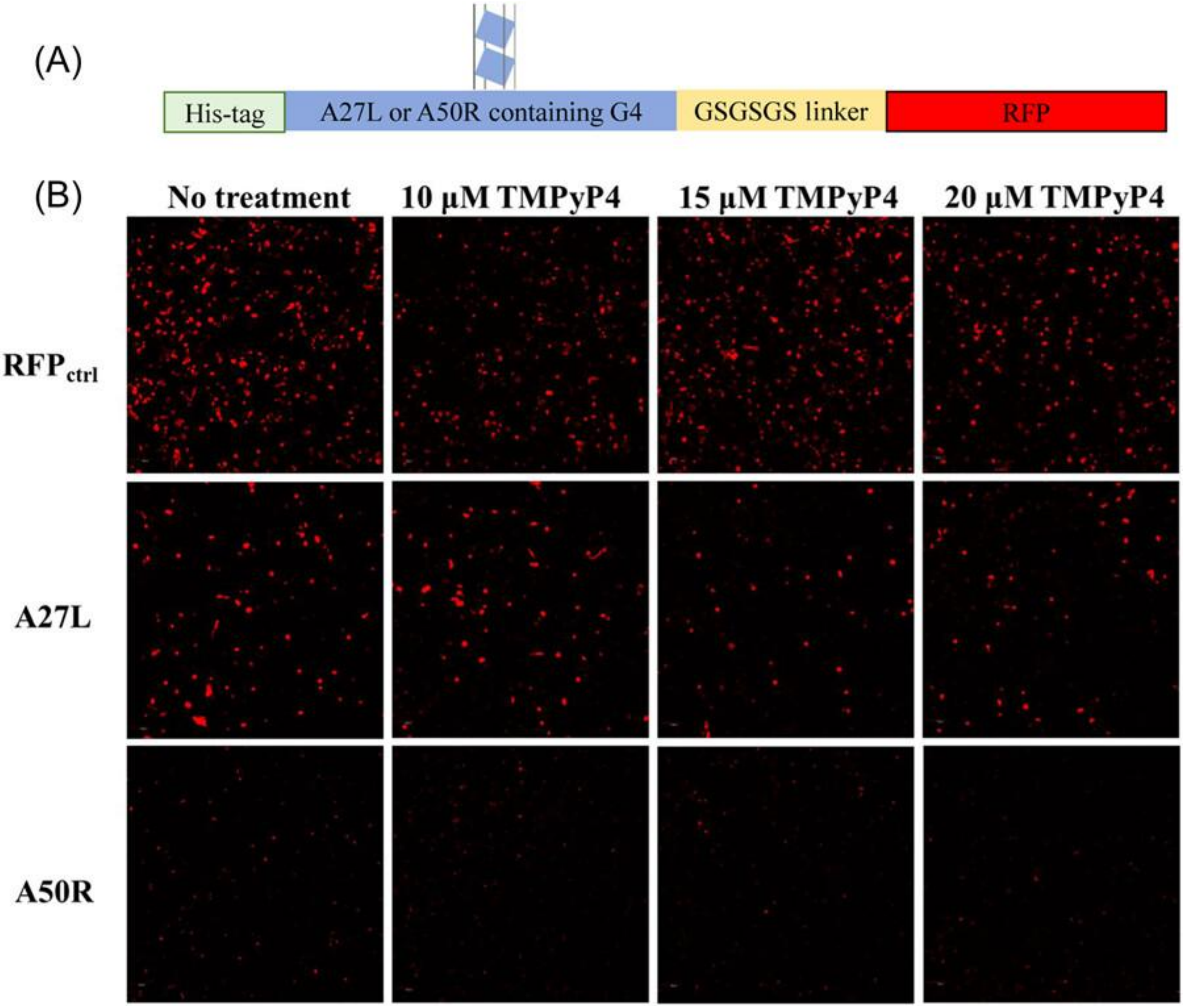


Figure 7.9 TMPyP4 treatment reduced expression of MKPX A27L and A50R genes in HEK293 cells. Cells transfected with RFP-expressing plasmid were treated with 0, 10, 15 and 20 µM of TMPyP4 (top panel) and used as control, while cells transfected with MPXV genes A27L and A50R, fusioned with RFP, were also treated with TMPyP4 and could be observed at middle and bottom panel, respectively. Porphyrin treatment clearly is inefficient in reducing RFP expression at the top panel. However, in A27L and A50R infected cells treated with 15 and 20 µM of TMPyP4, a substantial reduction in overall fluorescence could be noticed after stabilization of G4s when compared to their respective untreated cells. Cells were observed at confocal microscope using RFP excitation and emission at a 10x magnification.

7.5 Discussion

Structural insights and characteristics in the MPXV genome are poorly understood. Most of what is known were either achieved by sequence similarity with close-related Poxvirus or old literature reports that evaluate particular genomic organization features [46-48]. For the first time, a detailed mapping within more than 500 MPXV genomes from different outbreaks across the globe pinpointed various G-rich sequences that potentially fold into a G4 in solution. As the first step of bioinformatic screening, a genomic analysis using the Putative Hybrid-Quadruplex-forming sequence (PHQS) was performed. PHQS was primarily designed to determine the increasing prevalence and abundance of hybrid-quadruplex over intramolecular G4s by analyzing the transcription start site of several genomes [28]. However, we used the PHQS script solely to identify G-rich sequences throughout MPXV genomes regardless of inter- or intramolecular G4. An in-depth score-based prediction was subsequently used to classify putative G-rich sequences.

Collected putative G4-sequences ranged in nucleotide length from 15 to 48-mer (**Figure 7.1A**). However, G4-predictive scores were calculated by counting only the minimal sequence comprising the four G-tracts, considering that additional nucleotides could underestimate the score. QGRS Mapper uses retrieved nucleotide sequence information to generate a G-score based on loop size, considering the size and equivalent distribution of nucleotides on the loops [23]. The total number of tetrads also impacts the G-score; where more tetrads are formed, a more stable G4 structure can be formed. Even though the QGRS software does not provide a baseline score, a previously published paper established a threshold of 19 residues and demonstrated that all predicted oligos above the score folded into a G4 *in vitro* [15]. Alternatively, the G4RNA screener uses a machine learning approach combined with an experimentally validated G4 set to provide a sequence-based G-score using three objective analyses: cGcG, G4Hunter, and G4NNN [24]. cGcG ratio and G4Hunter score system are based on the competitive rates of Watson-Crick and Hoogsteen-based structures, where excessive cytosine over guanine can hinder G4 formation [24, 49]. Otherwise, G4NN uses an artificial neural network to evaluate similitudes among input sequences within an experimental G4 database [24]. Although both QGRS and G4RNAscreener were designed for RNA G4s, G4Hunter was modelled to analyze DNA G4s with high accuracy [50]. Also, the loop size and distribution, as well as the number of G-tracts for canonical G4s, are equivalent for DNA and RNA molecules, which does not impair the scoring system. Previous work had also predicted DNA G4s using the QGRS Mapper scoring approach [51].

Putative G4-sequences were pinpointed in different protein-coding regions of the MPXV genome, although only the Right terminal region possesses eight out of eleven sequences. MP1 was identified in the antisense strand of the A27L coding gene. A27L possess a reverse open-reading frame gene that encodes one of the viral A-type inclusion proteins frequently targeted for viral neutralization [52] and is responsible for environmental protection during *Orthopoxvirus* infections [53]. Meanwhile, MP2 was positioned inside the gene A50R that encodes the ATP-dependent DNA ligase, responsible for repairing nicked DNA duplexes in Vaccinia virus [54] with mutations associated in new lineages of MPXV isolates [55]. Furthermore, conservation studies analyzing disease-related *Orthopoxvirus* genomes against MPXV G-rich sequences showed that putative G4s are MPXV-exclusive (**Figure 7.1B**). The subtle variations observed in nucleotide arrangement impair the formation of G4 in other disease-related *Orthopoxviruses* once an essential G-tract is disrupted. Thus, the localization of G4 sequences in important coding sites and the strong conservation and exclusiveness collectively indicate that these structures may modulate either viral environmental protection or replication exclusively during MPXV infections.

CD spectroscopy has been established as a primary biophysical method to assess G4 folding in solution and to evaluate the topology of G4s by comparing different spectral profiles [39]. CD analysis of both MP1 and MP2 demonstrate their ability to fold in solution and assume a parallel topology compared to the mutant oligos (**Figure 7.2A**). Guinier plots also indicate that wild-type and mutants can fold in solution without aggregation (**Figure 7.3A**). Biophysical characterization using SAXS has been routinely used in our lab to characterize long non-coding RNA structures [41] but also to visualize small DNA structures that fold into quadruplexes. Using a 23-mer oligo, our group demonstrated structural differences in a wild-type Hepatitis B virus (HBV) quadruplex when single guanosine is replaced by adenines [14]. Interestingly, the Kratky plot (**Figure 7.3B**) and the pair-distribution function (**Figure 7.4**) are reasonably similar, with an almost bell-shaped curve when comparing wild-type quadruplexes. Additionally, MPXV G4s have a D_{max} of 44 and 45 Å, which is consistent with HBV quadruplex, whereas mutants' maximum distance is near 60 Å (**Table 7.2**). MP1 mutant showed a distinct profile with a slightly reduced D_{max}, 55 Å, exhibiting a slightly compacted structure but still significantly different than wild types. The three-dimensional low-resolution structures clearly represent the abovementioned differences and provide evidence of quadruplex folding in solution (**Figure 7.5**).

Biophysical characterization is consistent with MP1 and MP2 interaction with G4-binding partners. ThT was shown to serve as a fluorescence sensor for RNA [56] and DNA G4s [42] by stacking into G-tetrad and stabilizing the K^+ induced G4s. A considerable variation (about 20-30%) in the relative fluorescence among Telomeric and G4-forming oligos was noticed upon their interaction with ThT, which could be influenced by the number of tetrads (**Figure 7.6**). ThT stacks to the tetrads and the Telomeric G4 is formed by three tetrads, while MP oligos are formed by two. Also, the heightened fluorescence intensity for MP1 mutant indicates its ability to form one tetrad and therefore had higher than basal fluorescence counts, coherent with the somewhat compacted structure obtained by SAXS. DHX36$_{53\text{-}105}$ is an ATP-dependent helicase well-known to bind via the N-terminal region to G4 containing parallel topologies with high affinity and unwind them [14, 43, 44]. DHX36$_{53\text{-}105}$ -specific motif (DSM) localized in the N-terminal portion is essential in G4 recognition [57], and several factors like the number of G-tracts, stereochemistry and loop size have been shown to alter binding affinity [58]. MST studies have shown that MP1 and MP2 interact with the N-terminal region of DHX36$_{53\text{-}105}$in a comparable nanomolar range to previously reported DNA G4s structures (**Figure 7.7A**) [14, 57]. The difference in binding affinity between MP1 and MP2 could be explained by the difference in loop size and distribution; while MP1 has a smaller and most equally distributed loop, MP2 has a sizeable central loop and only one nucleotide in the side loops.

To investigate a potential therapeutic application of targeting MP1 and MP2 structures in a cellular system, interaction with TMPyP4 was also estimated. Besides targeting G4 structures in solution and *in vivo*, TMPyP4 has been described to stabilize DNA G4 structures and inhibit polymerase progression [20], stall ribosomes on RNA G_4C_2 repeats [45] and unfold an extremely stable RNA G4 present in mRNA to enhance protein translation [59]. **Figure 7.7B** demonstrates that MP1 and MP2 interact with TMPyP4 with sub-nanomolar affinities at a similar range compared to mutant oligos. Similarly observed in the Thioflavin T assay, the MP1 mutant interacted with TMPyP4, indicating its ability to fold into one solely tetrad and leading to the speculation that TMPyP4 preferentially associates with tetrads instead of loops in the G4 structure. In addition, it was previously revealed that TMPyP4 could recognize the DNA G4 structures simultaneously when associated with intrinsic cellular partners, such as DHX36$_{53-105}$ (**Figure 7.8**). Contrarily, DHX36$_{53-105}$ could not recognize or compete with G4 binding sites when TMPyP4 is previously associated. Previous literature reported TMPyP4 as an antiviral against Herpes Simplex Virus-1 [20], Zika [21] and Ebola virus [19] by stabilizing G4 structures and, consequently, inhibiting either transcription or translational events critical to viral replication. The porphyrin concentrations used in the aforementioned studies were similar to the range we used (**Figure 7.9**), and cell viability was equally assessed in all of them, demonstrating that concentrations above 25 µM could cause a significant reduction in cellular viability. TMPyP4 treatment demonstrated a remarkable effect in reducing MPXV gene expression without leading to complete inhibition of protein translation. A27L and A50R encode critical proteins for viral survival and protection and are frequently targeted for vaccine development in similar viruses. Consequently, the robust interaction within TMPyP4 and MPXV G4s suggests a promising therapeutic avenue to explore further in a context of a cellular viral infection.

7.6 Conclusion

In summary, our work provides a deep genomic-scale screening in MPXV isolates to identify G-rich sequences that potentially fold into G4 structures. Sequence similarities and conservation studies demonstrate that two putative G4s are conserved across various outbreak isolates and are MPXV-exclusive. A robust biophysical and biochemical characterization provides evidence that DNA G4s fold in solution, assuming a parallel topology. While this work has shown that the MPXV genome holds two DNA G4s that tightly interact with ThT and DHX36$_{53\text{-}105}$ *in vitro*, additional work is required to elucidate the biological significance of those structures using virology assays. This work also demonstrates an interaction between MPXV quadruplexes and the cationic porphyrin TMPyP4, and treatment with TMPyP4 substantially reduced gene expression of MPXV important protein. TMPyP4 need to be further validated as an alternative therapeutic approach in a context of a MPXV-infected cells to assess their antiviral activity. Several G4-binding chemicals have been explored within antiviral activity such as Braco-19 and Pyridostatin (PDS), which could be further explored in MPXV G4s. Finally, our work provide background to the further development of new therapeutics that target MPXV-specific structures.

7.7 References

1. Sklenovska, N. and M. Van Ranst, *Emergence of Monkeypox as the Most Important Orthopoxvirus Infection in Humans.* Front Public Health, 2018. **6**: p. 241.
2. Breman, J.G., et al., *Human monkeypox, 1970-79.* Bull World Health Organ, 1980. **58**(2): p. 165-82.
3. Bunge, E.M., et al., *The changing epidemiology of human monkeypox-A potential threat? A systematic review.* PLoS Negl Trop Dis, 2022. **16**(2): p. e0010141.
4. Reed, K.D., et al., *The detection of monkeypox in humans in the Western Hemisphere.* N Engl J Med, 2004. **350**(4): p. 342-50.
5. Kozlov, M., *Monkeypox goes global: why scientists are on alert.* Nature, 2022. **606**(7912): p. 15-16.
6. Alakunle, E.F. and M.I. Okeke, *Monkeypox virus: a neglected zoonotic pathogen spreads globally.* Nat Rev Microbiol, 2022. **20**(9): p. 507-508.
7. Alakunle, E., et al., *Monkeypox Virus in Nigeria: Infection Biology, Epidemiology, and Evolution.* Viruses, 2020. **12**(11).
8. Rimoin, A.W., et al., *Major increase in human monkeypox incidence 30 years after smallpox vaccination campaigns cease in the Democratic Republic of Congo.* Proc Natl Acad Sci U S A, 2010. **107**(37): p. 16262-7.
9. Dai, Y., et al., *A peculiar evolutionary feature of monkeypox virus.* bioRxiv, 2022: p. 2022.06.18.496696.
10. Spiegel, J., S. Adhikari, and S. Balasubramanian, *The Structure and Function of DNA G-quadruplexes.* Trends Chem, 2020. **2**(2): p. 123-136.
11. Balaratnam, S. and S. Basu, *Divalent cation-aided identification of physico-chemical properties of metal ions that stabilize RNA G-quadruplexes.* Biopolymers, 2015. **103**(7): p. 376-86.
12. Bhattacharyya, D., G. Mirihana Arachchilage, and S. Basu, *Metal Cations in G-quadruplex Folding and Stability.* Front Chem, 2016. **4**: p. 38.
13. Lombardi, E.P. and A. Londono-Vallejo, *A guide to computational methods for G-quadruplex prediction.* Nucleic Acids Res, 2020. **48**(3): p. 1603.
14. Meier-Stephenson, V., et al., *Identification and characterization of a G-quadruplex structure in the pre-core promoter region of hepatitis B virus covalently closed circular DNA.* J Biol Chem, 2021. **296**: p. 100589.
15. Holoubek, J., et al., *Guanine quadruplexes in the RNA genome of the tick-borne encephalitis virus: their role as a new antiviral target and in virus biology.* Nucleic Acids Res, 2022. **50**(8): p. 4574-4600.
16. Rhodes, D. and H.J. Lipps, *G-quadruplexes and their regulatory roles in biology.* Nucleic Acids Res, 2015. **43**(18): p. 8627-37.
17. Nakanishi, C. and H. Seimiya, *G-quadruplex in cancer biology and drug discovery.* Biochem Biophys Res Commun, 2020. **531**(1): p. 45-50.
18. Saranathan, N. and P. Vivekanandan, *G-quadruplexes: More Than Just a Kink in Microbial Genomes.* Trends Microbiol, 2019. **27**(2): p. 148-163.
19. Wang, S.R., et al., *Chemical Targeting of a G-quadruplex RNA in the Ebola Virus L Gene.* Cell Chem Biol, 2016. **23**(9): p. 1113-1122.

20. Artusi, S., et al., *Antiviral Activity of the G-quadruplex Ligand TMPyP4 against Herpes Simplex Virus-1.* Viruses, 2021. **13**(2).
21. Majee, P., et al., *Inhibition of Zika virus replication by G-quadruplex-binding ligands.* Mol Ther Nucleic Acids, 2021. **23**: p. 691-701.
22. Shammas, M.A., et al., *Telomerase inhibition and cell growth arrest by G-quadruplex interactive agent in multiple myeloma.* Mol Cancer Ther, 2003. **2**(9): p. 825-33.
23. Kikin, O., L. D'Antonio, and P.S. Bagga, *QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences.* Nucleic Acids Res, 2006. **34**(Web Server issue): p. W676-82.
24. Garant, J.M., J.P. Perreault, and M.S. Scott, *Motif independent identification of potential RNA G-quadruplexes by G4RNA screener.* Bioinformatics, 2017. **33**(22): p. 3532-3537.
25. Katoh, K. and D.M. Standley, *MAFFT multiple sequence alignment software version 7: improvements in performance and usability.* Mol Biol Evol, 2013. **30**(4): p. 772-80.
26. Nguyen, L.T., et al., *IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies.* Mol Biol Evol, 2015. **32**(1): p. 268-74.
27. Letunic, I. and P. Bork, *Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation.* Nucleic Acids Res, 2021. **49**(W1): p. W293-W296.
28. Xiao, S., et al., *Bioinformatic analysis reveals an evolutional selection for DNA:RNA hybrid G-quadruplex structures as putative transcription regulatory elements in warm-blooded animals.* Nucleic Acids Res, 2013. **41**(22): p. 10379-90.
29. Altschul, S.F., et al., *Basic local alignment search tool.* J Mol Biol, 1990. **215**(3): p. 403-10.
30. Sievers, F., et al., *Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.* Mol Syst Biol, 2011. **7**: p. 539.
31. Meier, M., et al., *Structure and hydrodynamics of a DNA G-quadruplex with a cytosine bulge.* Nucleic Acids Res, 2018. **46**(10): p. 5319-5331.
32. Franke, D., et al., *ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions.* J Appl Crystallogr, 2017. **50**(Pt 4): p. 1212-1225.
33. Panjkovich, A. and D.I. Svergun, *CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data.* Bioinformatics, 2018. **34**(11): p. 1944-1946.
34. Putnam, C.D., *Guinier peak analysis for visual and automated inspection of small-angle X-ray scattering data.* J Appl Crystallogr, 2016. **49**(Pt 5): p. 1412-1419.
35. Burke, J.E. and S.E. Butcher, *Nucleic acid structure characterization by small angle X-ray scattering (SAXS).* Curr Protoc Nucleic Acid Chem, 2012. **Chapter 7**: p. Unit7 18.
36. Svergun, D.I., *Determination of the regularization parameter in indirect-transform methods using perceptual criteria.* Journal of Applied Crystallography, 1992. **25**(4): p. 495-503.
37. Franke, D. and D.I. Svergun, *DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering.* J Appl Crystallogr, 2009. **42**(Pt 2): p. 342-346.
38. Volkov, V.V. and D.I. Svergun, *Uniqueness of ab initio shape determination in small-angle scattering.* Journal of Applied Crystallography, 2003. **36**(3-1): p. 860-864.

39. Del Villar-Guerra, R., J.O. Trent, and J.B. Chaires, *G-quadruplex Secondary Structure Obtained from Circular Dichroism Spectroscopy.* Angew Chem Int Ed Engl, 2018. **57**(24): p. 7171-7175.
40. Kim, D.N., et al., *Zinc-finger protein CNBP alters the 3-D structure of lncRNA Braveheart in solution.* Nat Commun, 2020. **11**(1): p. 148.
41. D'Souza, M.H., et al., *Biophysical characterisation of human LincRNA-p21 sense and antisense Alu inverted repeats.* Nucleic Acids Res, 2022. **50**(10): p. 5881-5898.
42. Mohanty, J., et al., *Thioflavin T as an efficient inducer and selective fluorescent sensor for the human telomeric G-quadruplex DNA.* J Am Chem Soc, 2013. **135**(1): p. 367-76.
43. Lattmann, S., et al., *Role of the amino terminal RHAU-specific motif in the recognition and resolution of guanine quadruplex-RNA by the DEAH-box RNA helicase RHAU.* Nucleic Acids Res, 2010. **38**(18): p. 6219-33.
44. Giri, B., et al., *G4 resolvase 1 tightly binds and unwinds unimolecular G4-DNA.* Nucleic Acids Res, 2011. **39**(16): p. 7161-78.
45. Mori, K., et al., *The porphyrin TMPyP4 inhibits elongation during the noncanonical translation of the FTLD/ALS-associated GGGGCC repeat in the C9orf72 gene.* J Biol Chem, 2021. **297**(4): p. 101120.
46. Mackett, M. and L.C. Archard, *Conservation and variation in Orthopoxvirus genome structure.* J Gen Virol, 1979. **45**(3): p. 683-701.
47. Shchelkunov, S.N., et al., *Analysis of the monkeypox virus genome.* Virology, 2002. **297**(2): p. 172-94.
48. Suraweera, C.D., et al., *Structural insight into tanapoxvirus-mediated inhibition of apoptosis.* FEBS J, 2020. **287**(17): p. 3733-3750.
49. Beaudoin, J.D., R. Jodoin, and J.P. Perreault, *New scoring system to identify RNA G-quadruplex folding.* Nucleic Acids Res, 2014. **42**(2): p. 1209-23.
50. Bedrat, A., L. Lacroix, and J.L. Mergny, *Re-evaluation of G-quadruplex propensity with G4Hunter.* Nucleic Acids Res, 2016. **44**(4): p. 1746-59.
51. Lago, S., et al., *The MDM2 inducible promoter folds into four-tetrad antiparallel G-quadruplexes targetable to fight malignant liposarcoma.* Nucleic Acids Res, 2021. **49**(2): p. 847-863.
52. Ahsendorf, H.P., et al., *Species-Specific Conservation of Linear Antigenic Sites on Vaccinia Virus A27 Protein Homologs of Orthopoxviruses.* Viruses, 2019. **11**(6).
53. Kastenmayer, R.J., et al., *Elimination of A-type inclusion formation enhances cowpox virus replication in mice: implications for orthopoxvirus evolution.* Virology, 2014. **452-453**: p. 59-66.
54. Moss, B., *Poxvirus DNA replication.* Cold Spring Harb Perspect Biol, 2013. **5**(9).
55. Abrahim, M., et al., *The emergence of new lineages of the Monkeypox virus could affect the 2022 outbreak.* bioRxiv, 2022: p. 2022.07.07.498743.
56. Xu, S., et al., *Thioflavin T as an efficient fluorescence sensor for selective recognition of RNA G-quadruplexes.* Sci Rep, 2016. **6**: p. 24793.
57. Chen, M.C., et al., *Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36.* Nature, 2018. **558**(7710): p. 465-469.

58. Mou, X. and C.K. Kwok, *Effect of RNA sequence context and stereochemistry on G-quadruplex-RHAU53 interaction.* Biochem Biophys Res Commun, 2020. **533**(4): p. 1135-1141.
59. Morris, M.J., et al., *The porphyrin TmPyP4 unfolds the extremely stable G-quadruplex in MT3-MMP mRNA and alleviates its repressive effect to enhance translation in eukaryotic cells.* Nucleic Acids Res, 2012. **40**(9): p. 4137-45.

8.0 Dissertation Conclusion

This thesis has made significant contributions to our understanding of viral DNA/RNA structures and their interactions with host proteins, specifically with ZIKV and MPXV viruses. This research:

1. Reviewed a collection of methods for modifying RNA for the capture of RNA-binding proteins, including small-molecule modification, the addition of aptamers, DNA-anchoring, and nucleotide substitution.
2. Identified all proteins that interacting with the 5’ and 3’ UTRs, directly and indirectly by performing an immunoprecipitation pulldown assay using the UTRs as bait.
3. Characterized the NCL structure and the ZIKV terminal regions by performing multi-angle/dynamic light scattering, circular dichroism spectroscopy, and small-angle X-ray scattering. Characterized the interaction of them using microscale thermophoresis and performed a siRNA knockdown of NCL in the Huh-7.5 hepatocarcinoma cells and showed for the first time that NCL is critical for the formation of functional virions.
4. Demonstrated that the ZIKV 3' UTR G4 is conserved across all ZIKV isolates and maintains its structure in a 3' TR full-length transcript, and demonstrate that the human DEAD-box helicases, DDX3X$_{132\text{-}607}$ and DDX17$_{135\text{-}555}$, bind to the 3' TR, and that DDX17$_{135\text{-}555}$ unfolds the G4 present in the 3' TR.
5. Identified two G4s within MPXV using CD spectroscopy and SAXS, show that a quadruplex binding small molecule with previously reported antiviral activity, TMPyP4, interacts with MPXV G4s with nanomolar affinity, and performed cell biology experiments suggesting that TMPyP4 treatment substantially reduced gene expression of MPXV proteins.

The methodologies employed in this research, particularly the RNA-protein interaction studies, align with cutting-edge techniques in the field. The use of RNA-baited immunoprecipitation pulldown assays builds a library of nucleic acid-protein interactions that are critical for understanding the binding landscape of genomes and transcripts, in our case, viral genomes.

Future Directions

Building on these findings and considering the evolving landscape of RNA-protein interaction studies, future research could:

1. Employ other advanced RNA labeling techniques, such as the Mango aptamer system, to further characterize ZIKV and Mpox RNA-protein interactions with enhanced sensitivity and specificity. Additionally, utilize CRISPR-based approaches, like the Csy4 system, for more stringent isolation of viral RNA-protein complexes, potentially revealing weaker or transient interactions.

2. Investigate the potential of identified RNA-protein interactions as targets for novel antiviral strategies, focusing on disrupting key viral replication processes.
3. Expand the comparative analysis to other emerging viruses, using the techniques and insights gained from this study to identify common mechanisms that could be targeted for broad-spectrum antiviral therapies.
4. Develop in vivo models to validate the findings and bridge the gap between molecular interactions and clinical manifestations of viral infections.

In conclusion, this thesis has significantly advanced our understanding of ZIKV and MPXV molecular virology, providing a solid foundation for future research. By incorporating state-of-the-art RNA-protein interaction methodologies, the work has not only contributed to the specific knowledge of these viruses but also to the broader field of viral DNA/RNA biology. The insights into these findings reveal new avenues of diagnostic and therapeutic developments that may provide better approaches against not only these viruses but also emerging viruses.