BIOPHYSICAL CHARACTERIZATION OF NON-CODING REGIONS OF VIRAL RNAS

Tyler Mrozowich
Bachelor of Science, University of Lethbridge, 2017

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

DOCTOR OF PHILOSOPHY

In

BIOMOLECULAR SCIENCE

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

BIOPHYSICAL CHARACTERIZATION OF NON-CODING REGIONS OF VIRAL RNAS

TYLER MROZOWICH

Date of Defence: 04, 20, 2022

Dr. Trushar Patel Thesis Supervisor	Associate Professor	Ph.D.
Dr. Steve Wiseman Thesis Examination Committee Member	Associate Professor	Ph.D.
Dr. Athanasios Zovoilis Thesis Examination Committee Member	Associate Professor	Ph.D.
Dr. Theresa Burg Internal External Examiner	Professor	Ph.D.
Dr. Olwyn Byron External Examiner	Professor	Ph.D.
Dr. Michael Gerken Chair, Thesis Examination Committee	Professor	Ph.D.

Dedication

This dissertation is dedicated to everyone who has believed in me and supported me throughout the years, especially my brother Ryan.

I think it might be time for that vacation to a lovely white sandy beach we have always talked about, Ry.

Abstract

Throughout the last few decades, it has emerged that the genomic organization of organisms is not simple, easily laid out, or straightforward. Long gone are the days of the central dogma, where DNA is transcribed to RNA, and RNA is translated to proteins. Non-coding RNA has emerged as an essential biological regulator of numerous cellular processes. These non-coding RNA are not simply a chain of ribonucleotides conferring function through the primary sequence; rather, functions of non-coding RNA are also related to the tertiary structure or 3-dimensional arrangement. This thesis focuses on the core theme of biophysical characterization of non-coding RNA structure and interactions with chapters and on four specific biophysical techniques I have used; multi-angle light scattering, microscale thermophoresis, small-angle X-ray scattering, and analytical ultracentrifugation. Using these techniques and others, we investigated non-coding RNA regions of different RNA-based viruses. We compared Murray Valley Encephalitis and Powassan virus', 5' and 3' TR atomic structures. We showed Japanese Encephalitis virus and Zika virus 5' TRs direct interaction with DDX3X human helicase and DDX17's direct interaction and unwinding of Rift Valley Fever virus' S-segment non-coding RNAs. Lastly, we characterized the cyclization interaction of the 5' and 3' TR of Japanese encephalitis using biophysical and computational methods. Taken together, this thesis demonstrates that many biophysical techniques can be used to investigate viral non-coding RNA structure and non-coding RNA interactions to uncover and validate meaningful biologically relevant interactions *in vitro*.

Contribution of Authors

This thesis contains multiple collaborative projects with numerous collaborators.

Chapter 3 focuses on MST and contains a manuscript titled "Microscale Thermophoresis: Warming up to a New Biomolecular Interaction Technique," I wrote in collaboration with Dr. Vanessa Meier-Stephenson and Dr. Trushar R Patel, published in The Biochemist Volume 41, Issue 2, Pages 8-12 on April 1, 2019. https://doi.org/10.1042/BIO04102008. This paper describes how MST works, the information obtained from MST experiments, its advantages and disadvantages, and the diagnosis of common problems. I wrote the manuscript with input from both Dr. Meier-Stephenson and Dr. Patel. This publication is reproduced with permission from Portland Press Ltd and the Creative Commons CC by license and re-formatted to fit thesis formatting.

Chapter 4 focuses on SAXS and contains a manuscript titled "Structural Studies of Macromolecules in Solution using Small Angle X-Ray Scattering," I wrote in collaboration with Steffane McLennan, Dr. Michael Overduin and Dr. Trushar R Patel, published in The Journal of Visualized Experiments, Volume 141, on November 5, 2018. https://dx.doi.org/10.3791/58538. This publication is unique because it is accompanied by a descriptive video (linked at the end of this chapter). The manuscript describes the basics of SAXS, a detailed description of data analysis following data collection and experimental results for nidogen-1 and laminin γ-1. I wrote the manuscript, along with input and data from Dr. Patel. Steffane McLennan and Dr. Overduin contributed to editing throughout. This publication is reproduced with permission from MyJoVE and the Creative Commons CC by license and re-formatted to fit thesis formatting.

Chapter 6 is a manuscript in collaboration with Amy Henrickson, Dr. Borries Demeler, and Dr. Trushar R Patel, published in *Viruses,* Volume 12, Issue 2 on February 8, 2020. https://doi.org/10.3390/v12020190. This publication focuses on characterizing the 5' and 3'

non-coding terminal regions of two flaviviruses, Murray Valley Encephalitis and Powassan Virus, through computational structure prediction and small angle x-ray scattering. The manuscript conception was both Dr. Patel and me. I wrote the manuscript with help and editing from the other authors. I performed and analyzed all experiments with help from Amy Henrickson and Dr. Demeler with the AUC portion. This publication is reproduced with permission from MDPI and the Creative Commons CC by license and re-formatted to fit thesis formatting.

Chapter 7 consists of a manuscript I wrote in collaboration with Corey R Nelson, whom I share co-authorship with, Sean Park, Simmone D'Souza, Amy Henrickson, Justin Vigar, Dr. Hans-Joachim Wieden, Dr. Raymon Owens, Dr. Borries Demeler and Dr. Trushar R Patel, published in The International Journal of Molecular Sciences, Volume 22, Issue 1 on December 23, 2020. https://doi.org/10.3390/ijms22010054. This publication focuses on the interaction between DDX17 human RNA helicase and Rift Valley Fever virus non-coding RNA. Corey and I were involved in all the experiments within the manuscript; however, he focused more on the protein aspect, and I focused on the RNA aspect. Sean Park helped perform experiments. The project conception was by Corey Nelson, Dr. Patel, and me, with some initial work done by Simmone D'Souza. Justin Vigar created the original RNA constructs, and Dr. Owens created the protein constructs. Corey and I contributed equally to writing the final draft with input from other authors. This publication is reproduced with permission from Corey Nelson, MDPI, and the Creative Commons CC by license and re-formatted to fit thesis formatting.

Chapter 8 consists of a manuscript I wrote in collaboration with Corey R Nelson (MSc 2021), whom I share co-authorship with, Sean Park, Darren Gemmill, and Dr. Trushar R Patel, published in *The International Journal of Molecular Sciences,* Volume 22, Issue 1 on January 2, 2021. https://doi.org/10.3390/ijms22010413. This publication focuses on the interaction of JEV and Zika 5' non-coding RNA terminal regions with DDX3X human RNA helicase. Corey and I were both involved in all the experiments manuscript, with him focusing more on the

protein aspect while I focused on the RNA aspect. Sean Park contributed to experiments, while Darren Gemmill provided the Zika 5' RNA. Corey and I both equally contributed to the writing of the final draft. This publication is reproduced with permission from Corey Nelson, MDPI, and the Creative Commons CC by license and re-formatted to fit thesis formatting.

This chapter consists of a manuscript written in collaboration with Sean Park, whom I share co-authorship with, Maria Waldl, Amy Henrickson, Scott Tersteeg, Corey Nelson, Anneke Deklerk, Dr. Borries Demeler, Dr. Ivo Hofacker, Dr. Michael Wolfinger, and Dr. Trushar R Patel. This manuscript was published in *Nucleic Acids Research* https://doi.org/10.1093/nar/gkad223 on March 31, 2023. This work focuses on the interaction of JEV 5' and 3' non-coding terminal regions using a combination of biophysical and computational methods. I performed all wet-lab experiments with assistance from Sean Park. Maria Waldl contributed the computational methods, and Corey Nelson assisted with MST data analysis. Scott Tersteeg and Anneke Deklerk helped produce mutant RNA for first-round revisions. Amy Henrickson and Dr. Demeler assisted with analytical ultracentrifugation. Sean additionally assisted with the writing of the final draft. The conception of the project was by me and Dr. Patel. This publication is reproduced with permission from Sean Park and the Creative Commons agreement (open access) and is re-formatted to fit thesis formatting.

Acknowledgments

First, thank you to all of the colleagues and collaborators who contributed to publications used within my thesis; your names will be found under each respective chapter. I joined the Patel lab six years ago when Dr. Patel joined the University of Lethbridge. I have seen colleagues come and go and numerous graduations and have thoroughly enjoyed my time here. I want to thank all of the members of the Patel Lab, both past and present. I have enjoyed meeting and getting to know each one of you. I would like to additionally thank the research labs in L9Q1, both past and present; the Demeler Lab, the Wieden Lab, and the Kothe Lab. I want to thank Dr. Hans-Joachim Wieden and Dr. Ute Kothe personally. Your constructive criticisms through many lab meetings inspired and directed me to be the scientist I am today. Thank you for not sugar-coating when I was wrong and pointing me in the right direction. Thank you to Dr. Steve Wiseman and Dr. Athan Zovoilis for being my committee members. Steve, thanks for always telling me I didn't know enough background information; it always pushed me to do better – Even If I didn't always listen.

Thank you to two specific members of Dr. Hans-Joachim Wieden's research lab, Dr. Luc Roberts (not for the investment advice, though) and Justin Vigar. You were both always around in your office for me to ask questions (often dumb ones). You both crafted the foundation of my ability to work within a research lab. I wouldn't be the scientist I am today without you both.

I would like to thank the lunch crew: Corey Nelson, Sean Park, and Michael D'Souza. You guys made it fun to come to work, even during the pandemic. I'm thankful I had such wonderful colleagues whom I can call friends. Thanks to my friend and office-mate Scott Tersteeg, you always brought a smile and a laugh to the office and made it easy to come to work each day.

To Dr. Trushar Patel, thank you for your guidance these last six years. If not for your Biochemistry 4000 course, where I met you, I would never have done an independent study, and therefore, I would not be where I am. Thank you for taking me to RiboWest 2017 in Vancouver; that experience made me realize that research was the career path that I wanted to pursue. Thank you for taking the chance on me and accepting me as an MSc student. Looking at my GPA, we both know you had no business taking that risk, but you never batted an eye. I hope I have made you proud throughout the last six years and will continue to do so. Thank you for the opportunities you have given me, like taking me to England for our SAXS beamtime, sending me to numerous conferences, and allowing me to be part of many collaborations, which have allowed me to transition to the next phase of my career in Boston. Thank you for being patient with me and always being positive, especially when I get discouraged. Thank you for all those beers we drank together; you have been not only a supervisor and mentor but also my friend.

To my family, I owe a wholehearted thank you—especially to my wonderful parents. I do not know how you supported me throughout all these years, even when I didn't deserve your support. I would not be where I am without your continuing love and support. Thanks to my brother Ryan – you always had my back when it came to anything and always believed me, especially when I've been a broke student for so long. Thanks for the countless hours of gaming; ever since we were kids, we'd fight, argue, and compete, but in the end, we're always on the same team, and I couldn't ask for a better teammate.

I want to thank the whole Henrickson family, ever since meeting everyone, you all have been nothing short of amazing. You have all accepted me as part of the family, and I genuinely feel appreciated.

Lastly, I would like to thank my wonderful and amazing fiancé Amy Henrickson. Amy - I wouldn't be where I am without you. Since you entered my life over three years ago, you have been the rock I have relied upon. You believed in me when I didn't believe in myself,

and I am forever grateful for that. You're the best thing that's ever happened to me, and I cannot wait for our life together – I know this is just the beginning. Thank you for your continuing support and love; you mean the world to me. Finally, to our wonderful dog Ellie, thanks for all the snuggles; you were the least helpful person in writing my thesis, and sometimes that's what I needed.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

2DSA	Two-dimensional spectrum analysis
A488	Alexa 488
A4F	Asymmetric-flow field-flow fractionation
ATP	Adenosine triphosphate
AUC	Analytical ultracentrifugation
BSA	Bovine serum albumin
CAPRIN	Cell cycle associated protein
cDNA	Copy DNA
CEM	Cryo-electron microscopy
cHP	Capsid coding region hairpin
CLIP-seq	Cross-linking immunoprecipitation sequency
CMF	Confocal max flux
CS	Cyclization sequence
DAR	Downstream AUG region
DB	Dumbbell
DENV	Dengue virus
DLS	Dynamic light scattering
D_{max}	Maximal particle dimension
dsRNA	Double stranded RNA
D_T	Diffusion coefficient
EDTA	Ethylenediaminetetraacetic acid
EOM	Ensemble optimization method
FFF	Field flow fractionation
FITC	Fluorescein isothiocyanate

f0	Frictional coefficient
FPLC	Fast protein liquid chromatography
FRET	Fluorescent resonance energy transfer
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
HEK	Human embryonic kidney
IEX-MALS	Ion exchange coupled multi-angle light scattering
IGR	Intergenic region
INF	interferon
ITC	Isothermal titration calorimetry
JEV	Japanese encephalitis virus
K_d	Dissociation constant
lncRNA	Long non-coding RNA
MALS	Multi Angle light scattering
mESC	Mouse embryonic stem cell
miRNA	Micro RNA
mRNA	Messenger RNA
MST	Microscale thermophoresis
MTase	Methyltransferase
MVEV	Murray Valley Encephalitis virus
M_W	Molecular weight
NCR	Non-coding region
ncRNA	Non-coding RNA
NMR	Nuclear magnetic resonance
NSD	Normalized spatial discrepancy
NTPase	Nucleotide triphosphatase
Nts	Nucleotides

OAS	2’ 5’ oligoadenylate synthetase
OPPF	Oxford protein production facility
ORF	Open reading frame
PDB	Protein data bank
PowV	Powassan virus
RdRp	RNA-dependent RNA-polymerase
R_g	Radius of gyration
RGD	Rayleight-Gans-Debye
R_h	Hydrodynamic radius
RI	Refractive index
RIG I	Retinoic acid-inducible gene I
RMS	Root mean square
RNAi	RNA interference
rpm	Revolutions per minute
rRNA	Ribosomal RNA
RTPase	Ribonucleotide triphosphatase
RVFV	Rift Valley Fever virus
SAXS	Small angle X-ray scattering
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SE-AUC	Sedimentation equilibrium AUC
SEC	Size exclusion chromatography
sfRNA	Subgenomic flaviviral RNA
sHP	Small hairpin
siRNA	Small interfering RNA
SL	Stem loop
SLA	Stem-loop A

SLB	Stem loop B
snRNA	Small nuclear RNA
snRNA	Small nuclear RNA
ssRNA	Single-stranded RNA
SLI	Stem-loop 1
SV-AUC	Sedimentation velocity AUC
T7 RNAp	T7 RNA polymerase
tRNA	Transfer RNA
TRs	Terminal regions
UAR	Upstream AUG region
UHPLC	Ultra high-performance liquid chromatography
UTR	Untranslated region
Vbar ($\bar{v}$)	Partial specific volume
WHO	World Health Organization
WNV	West Nile virus
χ^2	Chi squared
xg	Relative centrifugal force
XRC	X-Ray crystallography
YFV	Yellow fever virus
ZikV	Zika virus

Chapter 1. Introduction and Background

1.1 Non-coding RNA

Non-coding RNA (ncRNA) generally refers to RNA which does not encode for a protein; however, this is misleading, suggesting ncRNA does not have a function. It was generally accepted as the "central dogma" that DNA is transcribed to RNA, then that RNA is translated to protein (Figure 1.1). This notion was prevalent for decades, even with the discovery of tRNA (transfer RNA) and rRNA (ribosomal RNA) in the 1950s, which were ncRNA with biological roles (1). More discoveries occurred throughout the 1970s and '80s of functional ncRNAs, such as RNAse P (2), snRNAs (small nuclear RNA) (3), and 7SL (4). Even throughout the 1990s, significant ncRNA with biological functions like Xist was discovered, a ncRNA that regulates chromosome structure (5). It was not until the revelation that almost all the mammalian genome is, at some level, transcribed (6-8) and likely functional (9-12) which caused a shift in how scientists view ncRNA.

We now know that only 1-2% of the human genome codes for proteins (12), a stunning statistic when evaluating the amount of research dedicated to mRNA (messenger RNA) and its product. However, an explosion of research has widened our knowledge of ncRNA to the point where over 40 different RNA classes exist (13). These classes include a wide variety of RNA including but not limited to gRNA, group I and II introns, ribozymes, mRNA, miRNA (micro RNA), tRNA, riboswitches, rRNA, siRNA (small interfering RNA), snoRNA (small nucleolar RNA), snRNA (small nuclear RNA), and lncRNA (long non-coding RNA), with these RNAs implicated in numerous cellular processes, often as important regulators (14). A

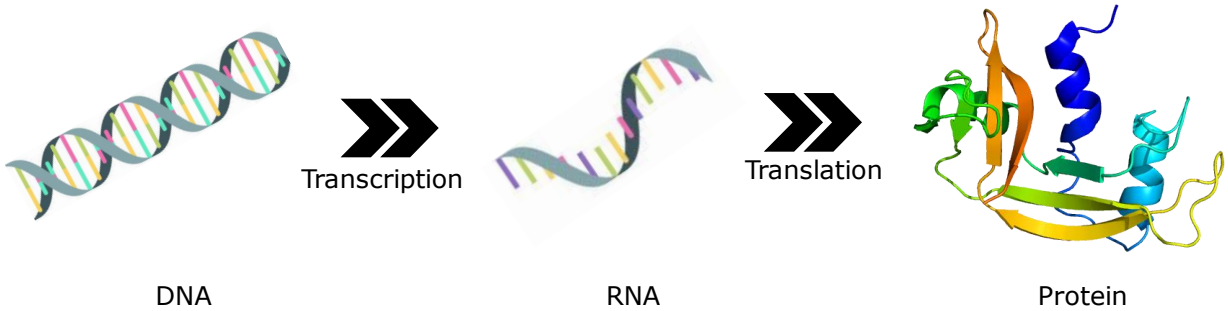


Figure 1.1 Central dogma of biology. DNA is transcribed to RNA which is then translated to proteins. Protein structure is PDB ID: 1kf5

comprehensive overview of each class of RNA and its effect on mammalian genetics would be impossible in this format; however, there are many excellent reviews on the subject (13,15-23). Therefore, for the remainder of this thesis, I will focus on lncRNA, specifically lncRNA in viral systems.

1.2 Long non-coding RNA

The accepted definition of lncRNA is an RNA transcript that is >200 nucleotides and not translated into a protein (24). If this definition seems broad, lacking, and ultimately unsatisfying, that is because it is. The need for such a blanket definition is because of the wide diversity of functions attributed to lncRNA. lncRNAs have been implicated in an enormous amount of biological processes such as neurodevelopment (21,25), transcription via repression of the human dihydrofolate gene (26), or through HOTAIR, which can repress transcription of the HOXD locus (27). LUST, another lncRNA can aid in elevating expression of the RBM5 splice variant (28), while MALAT1 regulates alternative splicing through modulating active serine/arginine splicing factors are examples of lncRNA in splicing (29). Translation has been shown to be affected by BC1 and BC200 effects on PABP1 (30). meiRNA is implicated in protein localization (31) along with Enod40 RNA (32). Cell structure integrity is implicated with the organization of cytokeratin filaments in the presence of VegT RNA in *Xenopus laevis* (33) and MEN epsilon/beta ncRNA is essential in the structural organization of paraspeckles (34). X-chromosome inactivation has been shown to be regulated by a lncRNA Xist (35-37). Cell survival and regulation of cell-cycle progression have been implicated with lncRNA PINC (38) and GAS5 (39). Apoptosis is affected by human cytomegalovirus through virally encoded beta2.7 lncRNA interactions (40) and lincRNA-p21 through p53-dependent transcriptional repression (41). AK028326 and AK141205 lncRNA are direct targets of Oct4 and Nanog, implicating ncRNA in mESC (mouse embryonic stem cell) pluripotency (42) and lincRNA-RoR implicated in stem cell reprogramming (43). B2 RNA has been shown to be

involved in heat shock response through interactions with RNA polymerase II (44,45). Cancer progression and other diseases have also been shown to have lncRNA association (46,47).

The long-standing biology trope that structure equals function is a widely held view in terms of organisms, tissues, cells, and proteins, but it is something that is likely also true of lncRNA. This structure-function relationship could help explain the relative lack of conservation of lncRNA sequences across organisms (48-50). Instead, it is argued that evolutionary selection acts upon structural domains in an "RNA modular code" hypothesis, with some evidence supporting this (51). While the biological functions of numerous lncRNA are extensive, and some are well characterized, there is a significant lack of knowledge regarding the structure of most lncRNAs. The cumulative number of RNA structures which have been "solved" compared with the number of protein structures is staggering (Figure 1.2). Seven times more protein structures are deposited into the PDB (protein data bank) (52) each year than cumulative RNA structures at the end of 2022. The last ten years of protein structure determination have seen, on average, 26 protein structures deposited each day into the PDB, which is a stark contrast to the deposition of a single RNA structure every five days (Figure 1.2). This wealth of protein structural knowledge has enabled researchers to find numerous well-defined and commonly observed fold-families and structural motifs (53). These motifs are sometimes very similar but nuanced enough and seen experimentally that they can be classified as different motifs such as β-hairpins, β-sandwiches, mixed β-sandwiches, β-barrels, and many more (53). These classified motifs have allowed researchers

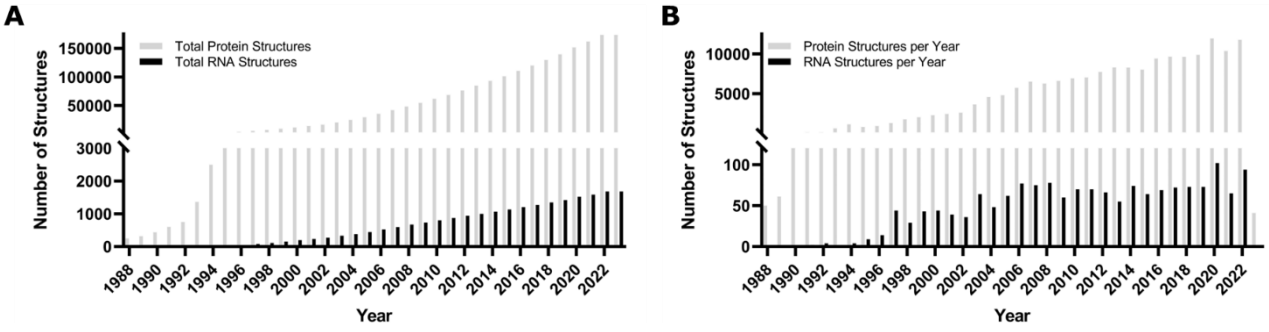


Figure 1.2. Statistical information on protein and RNA structures in the PDB. **A)** Cumulative number of structures deposited into the PDB **B)** Number of structures deposited each year into the PDB.

to assign commonly seen motifs to proteins with similar functions, allowing for easier structural determination and predictions. With the recent rise in machine learning, there is now the ability to train artificial intelligence to identify these structural motifs from primary sequences based on the thousands of known protein structures. This is evident in the rise of AlphaFold (54), which can accurately predict protein structure because of the depth of previous knowledge contained within the PDB. However, because of the number, or rather lack thereof RNA structures, and that most of these structures are <200 nucleotides in length, we do not have the same ability to predict RNA structure that we have for proteins. That is not to say that there is not considerable effort towards RNA structure determination and prediction. Numerous RNA secondary structure motifs have been identified and used to model predictions with the help of free energy minimization (55). This has allowed researchers to determine secondary structure from primary structure (56-64). However, predicting tertiary structure is considerably more complex and often only used on short RNA (65). This lack of RNA tertiary structure prediction comes from a lack of known RNA structures, and the lack of known RNA structures comes from decades of brushing off ncRNA as "junk RNA" not investing time or resources into characterizing RNA. Therefore, as we seek to explain RNA function as it pertains to structure, we need to investigate RNA structure from all angles, from the primary structure through quaternary, and from a foundational biophysical perspective.

1.3 Flaviviruses

Family *Flaviviridae* are small, positive-sense single-stranded RNA viruses that replicate within host cells of arthropods and/or vertebrates and include deadly viruses such as JEV (Japanese encephalitis virus), DENV (Dengue virus), ZikV (Zika virus), WNV (West Nile virus) and YFV (Yellow Fever virus). Due to the geographic diversity between flaviviruses (Figure 1.3), over half of the world's population is at risk for one or more flaviviral infections (66). For example, DENV now infects approximately 400 million individuals per year (67), YFV has encroached upon urban metropolises despite an effective vaccine (68), JEV presents an

ongoing public health risk and is emerging in different parts of the world, and modern medicine has not prevented the spread of WNV and Zika across the western hemisphere (69,70). The World Health Organization and the Centers for Disease Control both cite flaviviruses as a global health threat owing to the ease of transmission by mosquitoes as well as the lack of any efficient therapeutic or immunoprophylactic strategy (71). The flaviviral spread reflects their unique host vector, the mosquito, and the ability of these viruses to maintain stable populations of vertebrate animal reservoirs (72). Furthermore, global travel, changing environmental conditions, and poorly planned urbanization have resulted in a perfect storm of conditions to favor increasing flaviviral spread (73,74). This continued threat of flavivirus emergence and re-emergence emphasizes the need for continued research toward understanding flaviviral structure, biology, and biochemistry.

Flaviviruses are enclosed spherical virus particles encasing the flaviviral genome, a single-stranded (+) sense RNA molecule approximately 10-11k nucleotides long, depending on the species. The RNA genome contains a single open reading frame (ORF) flanked by non-coding terminal regions (TRs) (75) (Figure 1.4). The ORF codes for a single polypeptide which is cleaved by a combination of viral and host proteases (76); the resultant peptide chain produces three structural proteins and seven non-structural proteins (76). The three

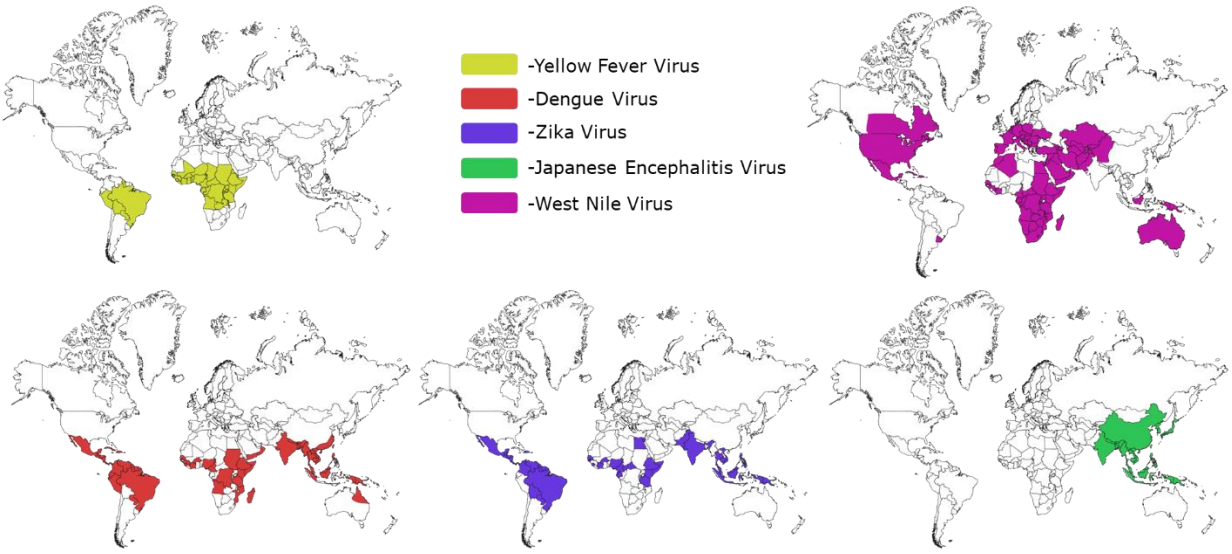


Figure 1.3 Global distribution of mosquito-born flaviviruses. Colour coded regions represent areas in which each virus is endemic.

structural proteins are C, prM, and E. The C protein forms the viral capsid but has numerous other functions, including interacting with host proteins and phosphorylation (77). The prM protein is a component of the flavivirus envelope and acts as a chaperone to help fold the E protein, the other envelope component (78). The E protein is responsible for viral entry and is an essential component of the viral envelope (78). The seven non-structural proteins are NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5. Non-structural protein 1 (NS1) is a secreted glycoprotein implicated in host immune-system evasion (79) and further facilitates viral replication through interacting with GAPDH (glyceraldehyde 3-phosphate dehydrogenase) and the uncleaved NS4A-2K-NS4B polyprotein (80). NS2A is a membrane protein that contributes to viral replication and assembly (81) by regulating the proteolytic cleavage of prM and E by interacting with the 3′UTR (untranslated region) (82). NS2B is a trans-membrane hydrophobic protein that can form an active serine protease with NS3 (83). Additionally, NS2B is implicated in inhibiting type 1 interferon production (84). Flaviviral NS3 is the second-largest viral protein at ~70 kDa playing multiple roles in replication (85). NS3 contains a DEAH-box helicase domain capable of NTPase (nucleotide triphosphatase), RTPase (ribonucleotide triphosphatase), (86) and dsRNA unwinding activities, which contributes to viral replication (87). NS4A plays an essential role in flaviviral infection through endoplasmic reticulum remodeling, which acts towards the replication complex formation along with NS2A, NS2B, and NS4B (88). The primary antiviral mediating flaviviral protein is NS4B, which inhibits nuclear localization of STAT1, consequently restricting IFN (interferon) α/β signaling (89) and

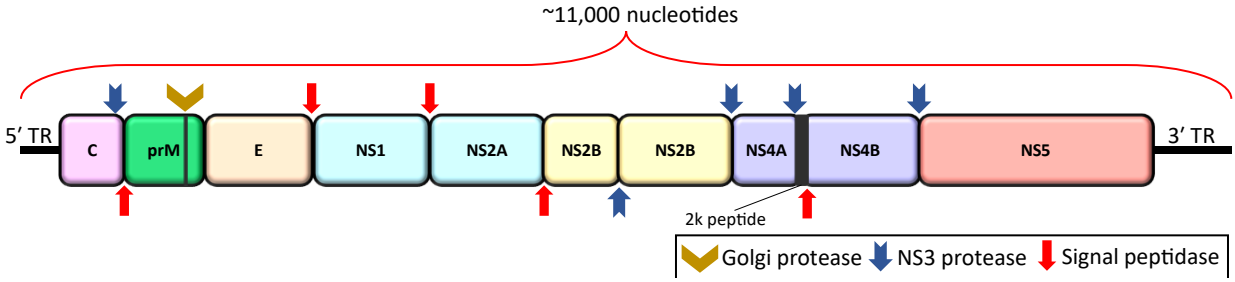


Figure 1.4. Flaviviral genomic arrangement. Three structural proteins and seven non-structural proteins are produced from a single polypeptide. Protein cleavage sites are marked with arrows.

RIG I (retinoic acid-inducible gene I) mediated activity (90). Viral NS5 is the largest viral protein and acts as the viral RNA-dependent RNA polymerase (91). NS5 is the most conserved flaviviral protein (92), containing both MTase (methyltransferase) (93) and RdRp (RNA-dependent RNA-polymerase) (94) domains. The NS5 MTase domain adds a type I cap (m^7GpppAmp) to the 5′ end of the (+) sense RNA transcript to evade 5′-3′ exoribonuclease activity while simultaneously avoiding host immune surveillance (95). Furthermore, it is hypothesized that viral genome cyclization is also required for the correct positioning of NS5 to replicate the viral genome (96).

1.4 Flavivirus non-coding RNA

The flanking structural TRs are highly conserved across all flaviviruses and appear as complex folding sequences essential for viral replication (97-99). Amongst flaviviruses, the 3′ end of the genome consists of a 400-800 nucleotide TR lacking a poly(A) tail and ending in a conserved CU_{OH} dinucleotide (75). This conserved CU_{OH} dinucleotide is implicated as a functional recognition site for the NS5 RNA-dependent RNA polymerase (100,101). The 5′ TR is considerably shorter, often ~100 nucleotides in length, and contains a type 1 cap structure N7 methylated guanosine (m^7GpppAmp) (102). An often-studied function of flaviviral terminal regions is the ability to cyclize, which has been shown computationally and experimentally in DENV and WNV (103-108). This cyclization is facilitated by an ~11 nucleotide complementary sequence found in the 5′ and 3′ terminal regions. The primary function of genomic cyclization in flaviviruses is to facilitate the replication of the genome through transcription of both the (+) and (-) sense RNA (108-111). There is experimental evidence that the NS5 MTase region interacts with the 5′ SLA (stem-loop A), which in conjunction with cyclization, brings the RdRp into proximity with the 3′ end of the genome to facilitate (-) sense RNA transcription (105,112,113). However, a question still lingers from those studies; does the NS5 protein remain bound to the 5′ SLA as it transcribes the (-) sense genome, or does the protein translocate to the 3′ end to facilitate (-) sense transcription? Interestingly, there is evidence

that there might be 5′ - 3′ base pairing interaction(s) between separate viral genomic RNA (114). However, this does not explain the difference between total (-) vs. (+) genomic RNA because this model would favor (-) sense transcription, but it is known that (-) sense transcripts are ~10x lower than (+) sense (115).

In addition to flaviviral terminal region importance in viral replication, flaviviruses have a distinct immune evasion mechanism shown to be important in stalling host XRN1 exoribonuclease. This XRN1 stalling results in the production of sfRNA (subgenomic flaviviral RNA) because of incomplete degradation of the 3′ TR (116-120). XRN1 stalling is only one function of sfRNA formation; the sfRNA itself has multiple cellular functions, including cytopathology (in cell culture) and viral pathogenesis in mice (117,121). Furthermore, sfRNA has been shown to interfere with RNAi (RNA interference) by acting as a decoy for RNAi mechanisms such as Dicer and Ago2, reducing cleavage of the viral dsRNA (122). Furthermore, sfRNA produced in mammalian cells has been shown to act against the RIG-I-dependent immune response in DENV infection and interferon innate immune responses in WNV infection (121,123). Another antiviral response of sfRNA is to act as binding sites for cellular translational proteins such as G3BP1, G3BP2, and CAPRIN (cell cycle associated protein), effectively sequestering them, reducing the amount of protein produced from interferon-induced mRNA (124).

Interestingly, when studying Kunjin viral infection in *Culex quinquefasciatus* mosquitos, it was shown that sfRNA only mildly suppressed the RNAi, which decreased viral transmission with little to no effect on mosquito fitness (125). Complementary to that previous study, it was shown that the generation of WNV sfRNA in mosquitos was necessary for infection and transmission by helping the virus bypass the midgut barrier, ensuring sufficient viral titer in the mosquito saliva (126). Both studies show the importance of the balance of sfRNA production in mosquito fitness, highlighted by subsequent studies showing excessive RNAi inhibition in DENV2 infection was lethal in mosquitos (127,128). A 2015 study

has put forth the hypothesis that the 3′TR in DENV is easily mutated to regulate the amount of sfRNA produced: less sfRNA in mosquitos, increasing upon human infection through 3′TR mutation (129). Furthermore, Kieft et al. proposed that a change in the tertiary RNA structure of the 3′ TR is what modulates the levels of sfRNA produced depending on either mammalian or insect host. A change in the tertiary structure of xrRNA1 and xrRNA2 (the primary sites of sfRNA production) in mosquito infection reduces the ability of xrRNA1 to stall Xrn1, causing a decrease in sfRNA production. This is further supported by experiments showing that when viral infection moves between mammalian and insect hosts, the virus accumulates mutations in its 3′ TR xrRNA2 structure, forming functionally critical tertiary interactions (130). Interestingly, the mutational changes occur almost exclusively in xrRNA2, which suggests that the virus maintains two xrRNA regions but modulates the structure of xrRNA1 by mutations made to xrRNA2, thus disrupting the tertiary structure.

Studies like Villordo et al., (130) and many others, show the importance of studying flaviviral terminal region structure; however, terminal regions and ncRNA are often considered less important than functional proteins. This bias suggests a systemic misunderstanding of the importance of ncRNA, where ncRNAs and ncRNA structure should be considered to be equally important as proteins and protein structure.

1.5 References

1. Hoagland, M.B., Stephenson, M.L., Scott, J.F., Hecht, L.I. and Zamecnik, P.C. (1958) A soluble ribonucleic acid intermediate in protein synthesis. *J Biol Chem*, **231**, 241-257.
2. Stark, B.C., Kole, R., Bowman, E.J. and Altman, S. (1978) Ribonuclease P: an enzyme with an essential RNA component. *Proceedings of the National Academy of Sciences of the United States of America*, **75**, 3717-3721.
3. Yang, V.W., Lerner, M.R., Steitz, J.A. and Flint, S.J. (1981) A small nuclear ribonucleoprotein is required for splicing of adenoviral early RNA sequences. *Proceedings of the National Academy of Sciences of the United States of America*, **78**, 1371-1375.
4. Walter, P. and Blobel, G. (1982) Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. *Nature*, **299**, 691-698.
5. Brockdorff, N., Ashworth, A., Kay, G.F., McCabe, V.M., Norris, D.P., Cooper, P.J., Swift, S. and Rastan, S. (1992) The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell*, **71**, 515-526.
6. Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C. *et al.* (2005) The transcriptional landscape of the mammalian genome. *Science*, **309**, 1559-1563.
7. Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigó, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E. *et al.* (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, **447**, 799-816.
8. Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F. *et al.* (2012) Landscape of transcription in human cells. *Nature*, **489**, 101-108.
9. Mattick, J.S., Taft, R.J. and Faulkner, G.J. (2010) A global view of genomic information--moving beyond the gene and the master regulator. *Trends Genet*, **26**, 21-28.
10. Ecker, J.R., Bickmore, W.A., Barroso, I., Pritchard, J.K., Gilad, Y. and Segal, E. (2012) Genomics: ENCODE explained. *Nature*, **489**, 52-55.
11. Pennisi, E. (2012) Genomics. ENCODE project writes eulogy for junk DNA. *Science*, **337**, 1159, 1161.
12. Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R. *et al.* (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57-74.
13. Cech, Thomas R. and Steitz, Joan A. (2014) The Non-coding RNA Revolution—Trashing Old Rules to Forge New Ones. *Cell*, **157**, 77-94.
14. Kaikkonen, M.U., Lam, M.T. and Glass, C.K. (2011) Non-coding RNAs as regulators of gene expression and epigenetics. *Cardiovasc Res*, **90**, 430-440.
15. Lee, Jeannie T. and Bartolomei, Marisa S. (2013) X-Inactivation, Imprinting, and Long Non-coding RNAs in Health and Disease. *Cell*, **152**, 1308-1323.
16. Batista, Pedro J. and Chang, Howard Y. (2013) Long Non-coding RNAs: Cellular Address Codes in Development and Disease. *Cell*, **152**, 1298-1307.
17. Carthew, R.W. and Sontheimer, E.J. (2009) Origins and Mechanisms of miRNAs and siRNAs. *Cell*, **136**, 642-655.
18. Ebert, Margaret S. and Sharp, Phillip A. (2012) Roles for MicroRNAs in Conferring Robustness to Biological Processes. *Cell*, **149**, 515-524.
19. Kopp, F. and Mendell, J.T. (2018) Functional Classification and Experimental Dissection of Long Non-coding RNAs. *Cell*, **172**, 393-407.

20. Ørom, Ulf A. and Shiekhattar, R. (2013) Long Non-coding RNAs Usher In a New Era in the Biology of Enhancers. *Cell*, **154**, 1190-1193.
21. Ponting, C.P., Oliver, P.L. and Reik, W. (2009) Evolution and Functions of Long Non-coding RNAs. *Cell*, **136**, 629-641.
22. Ulitsky, I. and Bartel, David P. (2013) lincRNAs: Genomics, Evolution, and Mechanisms. *Cell*, **154**, 26-46.
23. Statello, L., Guo, C.J., Chen, L.L. and Huarte, M. (2021) Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol*, **22**, 96-118.
24. Ma, L., Bajic, V.B. and Zhang, Z. (2013) On the classification of long non-coding RNAs. *RNA Biology*, **10**, 924-933.
25. Aliperti, V., Skonieczna, J. and Cerase, A. (2021) Long Non-Coding RNA (lncRNA) Roles in Cell Biology, Neurodevelopment and Neurological Disorders. *Non-coding RNA*, **7**.
26. Martianov, I., Ramadass, A., Serra Barros, A., Chow, N. and Akoulitchev, A. (2007) Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature*, **445**, 666-670.
27. Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E. *et al.* (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by non-coding RNAs. *Cell*, **129**, 1311-1323.
28. Rintala-Maki, N.D. and Sutherland, L.C. (2009) Identification and characterisation of a novel antisense non-coding RNA from the RBM5 gene locus. *Gene*, **445**, 7-16.
29. Tripathi, V., Ellis, J.D., Shen, Z., Song, D.Y., Pan, Q., Watt, A.T., Freier, S.M., Bennett, C.F., Sharma, A., Bubulya, P.A. *et al.* (2010) The nuclear-retained non-coding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell*, **39**, 925-938.
30. Muddashetty, R., Khanam, T., Kondrashov, A., Bundman, M., Iacoangeli, A., Kremerskothen, J., Duning, K., Barnekow, A., Hüttenhofer, A., Tiedge, H. *et al.* (2002) Poly(A)-binding protein is associated with neuronal BC1 and BC200 ribonucleoprotein particles. *J Mol Biol*, **321**, 433-445.
31. Watanabe, Y. and Yamamoto, M. (1994) S. pombe mei2+ encodes an RNA-binding protein essential for premeiotic DNA synthesis and meiosis I, which cooperates with a novel RNA species meiRNA. *Cell*, **78**, 487-498.
32. Campalans, A., Kondorosi, A. and Crespi, M. (2004) Enod40, a short open reading frame-containing mRNA, induces cytoplasmic localization of a nuclear RNA binding protein in Medicago truncatula. *Plant Cell*, **16**, 1047-1059.
33. Kloc, M., Wilk, K., Vargas, D., Shirato, Y., Bilinski, S. and Etkin, L.D. (2005) Potential structural role of non-coding and coding RNAs in the organization of the cytoskeleton at the vegetal cortex of Xenopus oocytes. *Development*, **132**, 3445-3457.
34. Sunwoo, H., Dinger, M.E., Wilusz, J.E., Amaral, P.P., Mattick, J.S. and Spector, D.L. (2009) MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res*, **19**, 347-359.
35. Brown, C.J., Ballabio, A., Rupert, J.L., Lafreniere, R.G., Grompe, M., Tonlorenzi, R. and Willard, H.F. (1991) A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature*, **349**, 38-44.
36. Brockdorff, N., Ashworth, A., Kay, G.F., Cooper, P., Smith, S., McCabe, V.M., Norris, D.P., Penny, G.D., Patel, D. and Rastan, S. (1991) Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. *Nature*, **351**, 329-331.
37. Lee, J.T. (2009) Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome. *Genes Dev*, **23**, 1831-1842.
38. Ginger, M.R., Shore, A.N., Contreras, A., Rijnkels, M., Miller, J., Gonzalez-Rimbau, M.F. and Rosen, J.M. (2006) A non-coding RNA is a potential marker of cell fate during

mammary gland development. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 5781-5786.
39. Mourtada-Maarabouni, M., Hedge, V.L., Kirkham, L., Farzaneh, F. and Williams, G.T. (2008) Growth arrest in human T-cells is controlled by the non-coding RNA growth-arrest-specific transcript 5 (GAS5). *J Cell Sci*, **121**, 939-946.
40. Reeves, M.B., Davies, A.A., McSharry, B.P., Wilkinson, G.W. and Sinclair, J.H. (2007) Complex I binding by a virally encoded RNA regulates mitochondria-induced cell death. *Science*, **316**, 1345-1348.
41. Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M.J., Kenzelmann-Broz, D., Khalil, A.M., Zuk, O., Amit, I., Rabani, M. *et al.* (2010) A large intergenic non-coding RNA induced by p53 mediates global gene repression in the p53 response. *Cell*, **142**, 409-419.
42. Sheik Mohamed, J., Gaughwin, P.M., Lim, B., Robson, P. and Lipovich, L. (2010) Conserved long non-coding RNAs transcriptionally regulated by Oct4 and Nanog modulate pluripotency in mouse embryonic stem cells. *Rna*, **16**, 324-337.
43. Loewer, S., Cabili, M.N., Guttman, M., Loh, Y.H., Thomas, K., Park, I.H., Garber, M., Curran, M., Onder, T., Agarwal, S. *et al.* (2010) Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat Genet*, **42**, 1113-1117.
44. Espinoza, C.A., Allen, T.A., Hieb, A.R., Kugel, J.F. and Goodrich, J.A. (2004) B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. *Nat Struct Mol Biol*, **11**, 822-829.
45. Zovoilis, A., Cifuentes-Rojas, C., Chu, H.P., Hernandez, A.J. and Lee, J.T. (2016) Destabilization of B2 RNA by EZH2 Activates the Stress Response. *Cell*, **167**, 1788-1802.e1713.
46. Tano, K. and Akimitsu, N. (2012) Long non-coding RNAs in cancer progression. *Frontiers in Genetics*, **3**.
47. Wapinski, O. and Chang, H.Y. (2011) Long non-coding RNAs and human disease. *Trends Cell Biol*, **21**, 354-361.
48. Wutz, A., Rasmussen, T.P. and Jaenisch, R. (2002) Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat Genet*, **30**, 167-174.
49. Tsai, M.C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E. and Chang, H.Y. (2010) Long non-coding RNA as modular scaffold of histone modification complexes. *Science*, **329**, 689-693.
50. Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P. *et al.* (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, **458**, 223-227.
51. Zhang, X., Rice, K., Wang, Y., Chen, W., Zhong, Y., Nakayama, Y., Zhou, Y. and Klibanski, A. (2010) Maternally expressed gene 3 (MEG3) non-coding ribonucleic acid: isoform structure, expression, and functions. *Endocrinology*, **151**, 939-947.
52. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The Protein Data Bank. *Nucleic acids research*, **28**, 235-242.
53. Sun, P.D., Foster, C.E. and Boyington, J.C. (2004) Overview of protein structural and functional folds. *Curr Protoc Protein Sci*, **Chapter 17**, Unit 17.11.
54. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A. *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583-589.
55. Moore, P.B. (1999) Structural Motifs in RNA. *Annual Review of Biochemistry*, **68**, 287-300.

56. Zhao, Q., Zhao, Z., Fan, X., Yuan, Z., Mao, Q. and Yao, Y. (2021) Review of machine learning methods for RNA secondary structure prediction. *PLoS Comput Biol*, **17**, e1009291.
57. Wayment-Steele, H.K., Kladwang, W., Strom, A.I., Lee, J., Treuille, A., Becka, A., Das, R. and Eterna, P. (2022) RNA secondary structure packages evaluated and improved by high-throughput experiments. *Nat Methods*, **19**, 1234-1242.
58. Xayaphoummine, A., Bucher, T. and Isambert, H. (2005) Kinefold web server for RNA/DNA folding path and structure prediction including pseudoknots and knots. *Nucleic acids research*, **33**, W605-610.
59. Xayaphoummine, A., Bucher, T., Thalmann, F. and Isambert, H. (2003) Prediction and statistics of pseudoknots in RNA structures using exactly clustered stochastic simulations. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 15310-15315.
60. Zuker, M. and Stiegler, P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic acids research*, **9**, 133-148.
61. Bompfünewerer, A.F., Backofen, R., Bernhart, S.H., Hertel, J., Hofacker, I.L., Stadler, P.F. and Will, S. (2008) Variations on RNA folding and alignment: lessons from Benasque. *Journal of Mathematical Biology*, **56**, 129-144.
62. Hofacker, I.L. and Stadler, P.F. (2006) Memory efficient folding algorithms for circular RNA secondary structures. *Bioinformatics*, **22**, 1172-1176.
63. Ding, Y. and Lawrence, C.E. (2003) A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic acids research*, **31**, 7280-7301.
64. Chan, C.Y., Lawrence, C.E. and Ding, Y. (2005) Structure clustering features on the Sfold Web server. *Bioinformatics*, **21**, 3926-3928.
65. Townshend, R.J.L., Eismann, S., Watkins, A.M., Rangan, R., Karelina, M., Das, R. and Dror, R.O. (2021) Geometric deep learning of RNA structure. *Science*, **373**, 1047-1051.
66. Du Pont, K.E., McCullagh, M. and Geiss, B.J. (2022) Conserved motifs in the flavivirus NS3 RNA helicase enzyme. *WIREs RNA*, **13**, e1688.
67. Bhatt, S., Gething, P.W., Brady, O.J., Messina, J.P., Farlow, A.W., Moyes, C.L., Drake, J.M., Brownstein, J.S., Hoen, A.G., Sankoh, O. *et al.* (2013) The global distribution and burden of dengue. *Nature*, **496**, 504-507.
68. Faria, N.R., Kraemer, M.U.G., Hill, S.C., Goes de Jesus, J., Aguiar, R.S., Iani, F.C.M., Xavier, J., Quick, J., du Plessis, L., Dellicour, S. *et al.* (2018) Genomic and epidemiological monitoring of yellow fever virus transmission potential. *Science*, **361**, 894-899.
69. Pierson, T.C. and Diamond, M.S. (2018) The emergence of Zika virus and its new clinical syndromes. *Nature*, **560**, 573-581.
70. Roehrig, J.T. (2013) West nile virus in the United States - a historical perspective. *Viruses*, **5**, 3088-3108.
71. Fernández-Sanlés, A., Ríos-Marco, P., Romero-López, C. and Berzal-Herranz, A. (2017) Functional Information Stored in the Conserved Structural RNA Domains of Flavivirus Genomes. *Frontiers in microbiology*, **8**, 546.
72. Pandit, P.S., Doyle, M.M., Smart, K.M., Young, C.C.W., Drape, G.W. and Johnson, C.K. (2018) Predicting wildlife reservoirs and global vulnerability to zoonotic Flaviviruses. *Nature Communications*, **9**, 5425.
73. Young, P.R. (2018) Arboviruses: A Family on the Move. *Adv Exp Med Biol*, **1062**, 1-10.
74. Tabachnick, W.J. (2016) Climate Change and the Arboviruses: Lessons from the Evolution of the Dengue and Yellow Fever Viruses. *Annu Rev Virol*, **3**, 125-145.
75. Wengler, G. and Wengler, G. (1981) Terminal sequences of the genome and replicatioe-form RNA of the flavivirus west nile virus: absence of poly(A) and possible role in RNA replication. *Virology*, **113**, 544-555.

76. Chambers, T.J., Hahn, C.S., Galler, R. and Rice, C.M. (1990) Flavivirus genome organization, expression, and replication. *Annual review of microbiology*, **44**, 649-688.
77. Zhang, X., Zhang, Y., Jia, R., Wang, M., Yin, Z. and Cheng, A. (2021) Structure and function of capsid protein in flavivirus infection and its applications in the development of vaccines and therapeutics. *Veterinary Research*, **52**, 98.
78. Dey, D., Poudyal, S., Rehman, A. and Hasan, S.S. (2021) Structural and biochemical insights into flavivirus proteins. *Virus Research*, **296**, 198343.
79. Scaturro, P., Cortese, M., Chatel-Chaix, L., Fischl, W. and Bartenschlager, R. (2015) Dengue Virus Non-structural Protein 1 Modulates Infectious Particle Production via Interaction with the Structural Proteins. *PLOS Pathogens*, **11**, e1005277.
80. Allonso, D., Andrade, I.S., Conde, J.N., Coelho, D.R., Rocha, D.C., da Silva, M.L., Ventura, G.T., Silva, E.M. and Mohana-Borges, R. (2015) Dengue Virus NS1 Protein Modulates Cellular Energy Metabolism by Increasing Glyceraldehyde-3-Phosphate Dehydrogenase Activity. *Journal of virology*, **89**, 11871-11883.
81. Wu, R.H., Tsai, M.H., Chao, D.Y. and Yueh, A. (2015) Scanning mutagenesis studies reveal a potential intramolecular interaction within the C-terminal half of dengue virus NS2A involved in viral RNA replication and virus assembly and secretion. *Journal of virology*, **89**, 4281-4295.
82. Xie, X., Zou, J., Zhang, X., Zhou, Y., Routh, A.L., Kang, C., Popov, V.L., Chen, X., Wang, Q.-Y., Dong, H. *et al.* (2019) Dengue NS2A Protein Orchestrates Virus Assembly. *Cell Host & Microbe*, **26**, 606-622.e608.
83. Li, X.D., Deng, C.L., Ye, H.Q., Zhang, H.L., Zhang, Q.Y., Chen, D.D., Zhang, P.T., Shi, P.Y., Yuan, Z.M. and Zhang, B. (2016) Transmembrane Domains of NS2B Contribute to both Viral RNA Replication and Particle Formation in Japanese Encephalitis Virus. *Journal of virology*, **90**, 5735-5749.
84. Aguirre, S., Luthra, P., Sanchez-Aparicio, M.T., Maestre, A.M., Patel, J., Lamothe, F., Fredericks, A.C., Tripathi, S., Zhu, T., Pintado-Silva, J. *et al.* (2017) Dengue virus NS2B protein targets cGAS for degradation and prevents mitochondrial DNA sensing during infection. *Nature Microbiology*, **2**, 17037.
85. Yusof, R., Clum, S., Wetzel, M., Murthy, H.M.K. and Padmanabhan, R. (2000) Purified NS2B/NS3 Serine Protease of Dengue Virus Type 2 Exhibits Cofactor NS2B Dependence for Cleavage of Substrates with Dibasic Amino Acids in Vitro*. *Journal of Biological Chemistry*, **275**, 9963-9969.
86. Wang, K., Zou, C., Wang, X., Huang, C., Feng, T., Pan, W., Wu, Q., Wang, P. and Dai, J. (2018) Interferon-stimulated TRIM69 interrupts dengue virus replication by ubiquitinating viral non-structural protein 3. *PLOS Pathogens*, **14**, e1007287.
87. Luo, D., Xu, T., Hunke, C., Grüber, G., Vasudevan, S.G. and Lescar, J. (2008) Crystal structure of the NS3 protease-helicase from dengue virus. *Journal of virology*, **82**, 173-183.
88. Hung, Y.-F., Schwarten, M., Hoffmann, S., Willbold, D., Sklan, E.H. and Koenig, B.W. (2015) Amino Terminal Region of Dengue Virus NS4A Cytosolic Domain Binds to Highly Curved Liposomes. *Viruses*, **7**, 4119-4130.
89. Muñoz-Jordán, J.L., Sánchez-Burgos, G.G., Laurent-Rolle, M. and García-Sastre, A. (2003) Inhibition of interferon signaling by dengue virus. *Proceedings of the National Academy of Sciences*, **100**, 14333-14338.
90. Ishikawa, H., Ma, Z. and Barber, G.N. (2009) STING regulates intracellular DNA-mediated, type I interferon-dependent innate immunity. *Nature*, **461**, 788-792.
91. Westaway, E.G. (1987) Flavivirus replication strategy. *Advances in virus research*, **33**, 45-90.
92. Tay, M.Y.F., Smith, K., Ng, I.H.W., Chan, K.W.K., Zhao, Y., Ooi, E.E., Lescar, J., Luo, D., Jans, D.A., Forwood, J.K. *et al.* (2016) The C-terminal 18 Amino Acid Region of Dengue Virus NS5 Regulates its Subcellular Localization and Contains a Conserved

Arginine Residue Essential for Infectious Virus Production. *PLOS Pathogens*, **12**, e1005886.
93. Zhao, Y., Soh, T.S., Lim, S.P., Chung, K.Y., Swaminathan, K., Vasudevan, S.G., Shi, P.-Y., Lescar, J. and Luo, D. (2015) Molecular basis for specific viral RNA recognition and 2′-O-ribose methylation by the dengue virus non-structural protein 5 (NS5). *Proceedings of the National Academy of Sciences*, **112**, 14834-14839.
94. Yap, T.L., Xu, T., Chen, Y.L., Malet, H., Egloff, M.P., Canard, B., Vasudevan, S.G. and Lescar, J. (2007) Crystal structure of the dengue virus RNA-dependent RNA polymerase catalytic domain at 1.85-angstrom resolution. *Journal of virology*, **81**, 4753-4765.
95. Daffis, S., Szretter, K.J., Schriewer, J., Li, J., Youn, S., Errett, J., Lin, T.-Y., Schneller, S., Zust, R., Dong, H. *et al.* (2010) 2′-O methylation of the viral mRNA cap evades host restriction by IFIT family members. *Nature*, **468**, 452-456.
96. Villordo, S.M. and Gamarnik, A.V. (2009) Genome cyclization as strategy for flavivirus RNA replication. *Virus Res*, **139**, 230-239.
97. Brinton, M.A. (2013) Replication cycle and molecular biology of the West Nile virus. *Viruses*, **6**, 13-53.
98. Brinton, M.A. and Basu, M. (2015) Functions of the 3' and 5' genome RNA regions of members of the genus Flavivirus. *Virus Res*, **206**, 108-119.
99. Brinton, M.A. and Dispoto, J.H. (1988) Sequence and secondary structure analysis of the 5'-terminal region of flavivirus genome RNA. *Virology*, **162**, 290-299.
100. Khromykh, A.A., Kondratieva, N., Sgro, J.Y., Palmenberg, A. and Westaway, E.G. (2003) Significance in replication of the terminal nucleotides of the flavivirus genome. *Journal of virology*, **77**, 10623-10629.
101. Nomaguchi, M., Ackermann, M., Yon, C., You, S. and Padmanabhan, R. (2003) De novo synthesis of negative-strand RNA by Dengue virus RNA-dependent RNA polymerase in vitro: nucleotide, primer, and template parameters. *Journal of virology*, **77**, 8831-8842.
102. Ray, D., Shah, A., Tilgner, M., Guo, Y., Zhao, Y., Dong, H., Deas, T.S., Zhou, Y., Li, H. and Shi, P.Y. (2006) West Nile virus 5'-cap structure is formed by sequential guanine N-7 and ribose 2'-O methylations by non-structural protein 5. *Journal of virology*, **80**, 8362-8370.
103. Alvarez, D.E., Lodeiro, M.F., Ludueña, S.J., Pietrasanta, L.I. and Gamarnik, A.V. (2005) Long-range RNA-RNA interactions circularize the dengue virus genome. *Journal of virology*, **79**, 6631-6643.
104. Alvarez, D.E., Filomatori, C.V. and Gamarnik, A.V. (2008) Functional analysis of dengue virus cyclization sequences located at the 5' and 3'UTRs. *Virology*, **375**, 223-235.
105. Dong, H., Zhang, B. and Shi, P.Y. (2008) Terminal structures of West Nile virus genomic RNA and their interactions with viral NS5 protein. *Virology*, **381**, 123-135.
106. Polacek, C., Foley, J.E. and Harris, E. (2009) Conformational changes in the solution structure of the dengue virus 5' end in the presence and absence of the 3' untranslated region. *Journal of virology*, **83**, 1161-1166.
107. Zhang, B., Dong, H., Stein, D.A., Iversen, P.L. and Shi, P.-Y. (2008) West Nile virus genome cyclization and RNA replication require two pairs of long-distance RNA interactions. *Virology*, **373**, 1-13.
108. Corver, J., Lenches, E., Smith, K., Robison, R.A., Sando, T., Strauss, E.G. and Strauss, J.H. (2003) Fine mapping of a cis-acting sequence element in yellow fever virus RNA that is required for RNA replication and cyclization. *Journal of virology*, **77**, 2265-2270.
109. Khromykh, A.A., Meka, H., Guyatt, K.J. and Westaway, E.G. (2001) Essential role of cyclization sequences in flavivirus RNA replication. *Journal of virology*, **75**, 6719-6728.
110. Lo, M.K., Tilgner, M., Bernard, K.A. and Shi, P.Y. (2003) Functional analysis of mosquito-borne flavivirus conserved sequence elements within 3' untranslated region

of West Nile virus by use of a reporting replicon that differentiates between viral translation and RNA replication. *Journal of virology*, **77**, 10004-10014.
111. Alvarez, D.E., De Lella Ezcurra, A.L., Fucito, S. and Gamarnik, A.V. (2005) Role of RNA structures present at the 3'UTR of dengue virus on translation, RNA synthesis, and viral replication. *Virology*, **339**, 200-212.
112. Filomatori, C.V., Lodeiro, M.F., Alvarez, D.E., Samsa, M.M., Pietrasanta, L. and Gamarnik, A.V. (2006) A 5' RNA element promotes dengue virus RNA synthesis on a circular genome. *Genes Dev*, **20**, 2238-2249.
113. Lodeiro, M.F., Filomatori, C.V. and Gamarnik, A.V. (2009) Structural and functional studies of the promoter element for dengue virus RNA replication. *Journal of virology*, **83**, 993-1008.
114. Lott, W.B. and Doran, M.R. (2013) Do RNA viruses require genome cyclisation for replication? *Trends Biochem Sci*, **38**, 350-355.
115. Chu, P.W.G. and Westaway, E.G. (1985) Replication strategy of Kunjin Virus: Evidence for recycling role of replicative form RNA as template in semiconservative and asymmetric replication. *Virology*, **140**, 68-79.
116. Urosevic, N., van Maanen, M., Mansfield, J.P., Mackenzie, J.S. and Shellam, G.R. (1997) Molecular characterization of virus-specific RNA produced in the brains of flavivirus-susceptible and -resistant mice after challenge with Murray Valley encephalitis virus. *J Gen Virol*, **78 (Pt 1)**, 23-29.
117. Pijlman, G.P., Funk, A., Kondratieva, N., Leung, J., Torres, S., van der Aa, L., Liu, W.J., Palmenberg, A.C., Shi, P.Y., Hall, R.A. *et al.* (2008) A highly structured, nuclease-resistant, non-coding RNA produced by flaviviruses is required for pathogenicity. *Cell Host Microbe*, **4**, 579-591.
118. Silva, P.A., Pereira, C.F., Dalebout, T.J., Spaan, W.J. and Bredenbeek, P.J. (2010) An RNA pseudoknot is required for production of yellow fever virus subgenomic RNA by the host nuclease XRN1. *Journal of virology*, **84**, 11395-11406.
119. Chapman, E.G., Costantino, D.A., Rabe, J.L., Moon, S.L., Wilusz, J., Nix, J.C. and Kieft, J.S. (2014) The structural basis of pathogenic subgenomic flavivirus RNA (sfRNA) production. *Science*, **344**, 307-310.
120. Akiyama, B.M., Laurence, H.M., Massey, A.R., Costantino, D.A., Xie, X., Yang, Y., Shi, P.Y., Nix, J.C., Beckham, J.D. and Kieft, J.S. (2016) Zika virus produces non-coding RNAs using a multi-pseudoknot structure that confounds a cellular exonuclease. *Science*, **354**, 1148-1152.
121. Schuessler, A., Funk, A., Lazear, H.M., Cooper, D.A., Torres, S., Daffis, S., Jha, B.K., Kumagai, Y., Takeuchi, O., Hertzog, P. *et al.* (2012) West Nile virus non-coding subgenomic RNA contributes to viral evasion of the type I interferon-mediated antiviral response. *Journal of virology*, **86**, 5708-5718.
122. Moon, S.L., Anderson, J.R., Kumagai, Y., Wilusz, C.J., Akira, S., Khromykh, A.A. and Wilusz, J. (2012) A non-coding RNA produced by arthropod-borne flaviviruses inhibits the cellular exoribonuclease XRN1 and alters host mRNA stability. *Rna*, **18**, 2029-2040.
123. Manokaran, G., Finol, E., Wang, C., Gunaratne, J., Bahl, J., Ong, E.Z., Tan, H.C., Sessions, O.M., Ward, A.M., Gubler, D.J. *et al.* (2015) Dengue subgenomic RNA binds TRIM25 to inhibit interferon expression for epidemiological fitness. *Science*, **350**, 217-221.
124. Bidet, K., Dadlani, D. and Garcia-Blanco, M.A. (2014) G3BP1, G3BP2 and CAPRIN1 Are Required for Translation of Interferon Stimulated mRNAs and Are Targeted by a Dengue Virus Non-coding RNA. *PLOS Pathogens*, **10**, e1004242.
125. Moon, S.L., Dodd, B.J., Brackney, D.E., Wilusz, C.J., Ebel, G.D. and Wilusz, J. (2015) Flavivirus sfRNA suppresses antiviral RNA interference in cultured cells and mosquitoes and directly interacts with the RNAi machinery. *Virology*, **485**, 322-329.

126. Göertz, G.P., Fros, J.J., Miesen, P., Vogels, C.B.F., van der Bent, M.L., Geertsema, C., Koenraadt, C.J.M., van Rij, R.P., van Oers, M.M. and Pijlman, G.P. (2016) Noncoding Subgenomic Flavivirus RNA Is Processed by the Mosquito RNA Interference Machinery and Determines West Nile Virus Transmission by Culex pipiens Mosquitoes. *Journal of virology*, **90**, 10145-10159.
127. Cirimotich, C.M., Scott, J.C., Phillips, A.T., Geiss, B.J. and Olson, K.E. (2009) Suppression of RNA interference increases alphavirus replication and virus-associated mortality in Aedes aegypti mosquitoes. *BMC Microbiol*, **9**, 49.
128. Khoo, C.C., Doty, J.B., Heersink, M.S., Olson, K.E. and Franz, A.W. (2013) Transgene-mediated suppression of the RNA interference pathway in Aedes aegypti interferes with gene silencing and enhances Sindbis virus and dengue virus type 2 replication. *Insect Mol Biol*, **22**, 104-114.
129. Kieft, J.S., Rabe, J.L. and Chapman, E.G. (2015) New hypotheses derived from the structure of a flaviviral Xrn1-resistant RNA: Conservation, folding, and host adaptation. *RNA Biol*, **12**, 1169-1177.
130. Villordo, S.M., Filomatori, C.V., Sánchez-Vargas, I., Blair, C.D. and Gamarnik, A.V. (2015) Dengue virus RNA structure specialization facilitates host adaptation. *PLoS Pathog*, **11**, e1004604.

Chapter 2. Multi-Angle and Dynamic Light Scattering

2.1 Foreword

Chapters 2-5 are written because of my extensive focus on learning and applying numerous biophysical techniques throughout my Ph.D. While this is not an exhaustive list, I primarily focused on four techniques, microscale thermophoresis (MST), MALS (multi-angle light scattering), small-angle X-ray scattering (SAXS), and analytical ultracentrifugation (AUC). This chapter focuses on multi-angle light scattering; specifically, size-exclusion coupled MALS, the theory behind the technique, the practical application, how to perform the technique, and technical considerations.

2.2 Abstract

Biomolecular characterization is an essential part of *in vitro* interaction experiments and structural studies. Often, biomolecules used for *in vitro* experiments are recombinantly expressed in biological species like Escherichia coli (prokaryotic expression system), a eukaryotic system like *Saccharomyces cerevisiae* (budding yeast), insect expression system like *Spodoptera frugiperda* (fall armyworm), or a mammalian expression system like HeLA cells, or HEK 293 (human embryonic kidney). Alternatively, nucleic acid biomolecules are produced enzymatically outside of a cell via *in vitro* transcription reactions. Regardless of how a biomolecule is produced, it needs to be assessed for biophysical properties. These often include size and shape (which can be combined to determine structure), stability, oligomerization, and purity. Therefore, we need techniques to probe the biophysical characteristics of these recombinantly expressed and purified biomolecules. Here, I will describe multi-angle light scattering, a technique used to accurately determine the molar mass of molecules in solution.

2.3 Theory

During the 19th century, John William Strutt, also known as Lord Rayleigh, offered the first explanation of why the sky is blue on a clear day. He based his explanation upon previous

equations describing light and its interactions with matter by James Clerk Maxwell in 1865. This description of the blue sky evolved into Rayleigh's theory describing the scattering of light by large macromolecules in solution, which later became known as the Rayleigh-Gans-Debye (RGD) theory of light scattering.

In basic terms, a solution of macro or nanoparticles is illuminated by a polarized beam of light (laser). An electric field is produced perpendicular to the plane, where angular and intensity-dependent scattered light can be measured. As a convention, the polarized direction is coined "vertical," while the measurement or scattering direction is referred to as "horizontal." The angular dependence within the horizontal plane is where information about the size of the macromolecule is derived, whereas the intensity dependence is where the molar mass information can be derived.

The theoretical considerations for the technique were condensed by Bruno Zimm (1), who used equations from Einstein (2), Rayleigh (3), Debye (4), and others (5-7), leading to the following equation:

$$\frac{K^* c}{R\,(\theta, c)} = \frac{1}{M_w P(\theta)} + 2A_2 c \qquad (2.1)$$

In this expression, $R\,(\Theta,c)$ is a function of the scattering angle Θ and concentration c. It is directly proportionate to the intensity of scattered light, irrespective of light scattered by solvent alone. Solute concentration is represented by c, with M_w representing the average molecular weight of the solute in solution. A_2 is representative of the second virial coefficient, a term dealing with osmotic pressure. $P\,(\Theta)$ is a function of the angular dependence of scattered light, which can be related to the RMS (mass averaged root mean square) radius, also known as R_g (radius of gyration). K^* is a constant, represented by equation 2:

$$K^* = 4\pi^2 \left(\frac{dn}{dc}\right)^2 \frac{{n_0}^2}{N_A {\lambda_0}^4}$$

(1.2)

In equation 2.2, the *dn/dc* term represents the specific refractive index increment, which is a change in the refractive index (*dn*) of a solution with respect to a change in its solute concentration (*dc*). N_a is Avagadro's number, which needs to be included when concentration is measured in *g/mL* and molar mass in *g/mol*. The refraction index of the solvent is denoted as n_0, with λ_0 representing the wavelength of the laser (in a vacuum). Equation 2.2 requires two things to be calculated or known beforehand: the *dn/dc* of the solute in solution, and the n_0, the refractive index of the solvent. Often, *dn/dc* is considered a constant based on the kind of biomolecule under investigation; for example, nucleic acids are assumed to have a *dn/dc* of 0.1720 mL/g (8), and proteins are assumed to have a *dn/dc* of 0.1850 mL/g (9). Second, the refractive index of the solvent n_0 must also be calculated or measured experimentally.

Modern MALS instruments often have an in-line refractometer that calculates the refractive index in the same volume as the scattering experiment. In addition to these two values, concentration of the solute in the solution needs to be calculated. Concentration calculation can be done in two ways: the first is to calculate the absorbance and then use

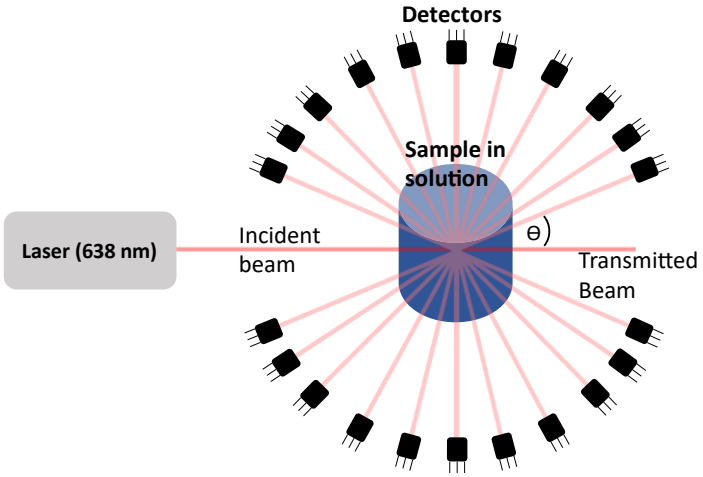


Figure 2.1. Schematic representation of a standard multi-angle light scattering setup.

Beer/Lambert law along with the extinction coefficient (10). The second way is to calculate concentration based on the change in the refractive index (11), measuring the difference between the refractive index of the solvent versus the solute and solvent in the solution (12). While a single angle is enough to perform a molecular weight measurement, modern light scattering detectors use an array of detectors to collect intensity information at multiple angles (Figure 2.1). Using multiple angles reduces the noise and increases the certainty of the y-intercept, reducing the error in the molar mass (13).

Additionally, modern MALS apparatuses are often connected in line with size exclusion chromatography (SEC) or field flow fractionation (FFF). Adding one of these techniques means that solute molecules are fractionated before light scattering. The most significant benefit this incurs over batch MALS is the determination of the molecular weight of a heterogenous solution. Both SEC and FFF will separate molecules based on size and shape, resulting in the ability to calculate the molecular mass of a gaussian distribution across a given timeframe or elution volume (13). SEC is the most often used because it is simple and easily performed by numerous different chromatography machines. However, SEC suffers from an underlying problem: choosing the appropriate column resolution. Typically, SEC resin resolves across a given molecular weight range; for example, a Superdex 200 increase™ (Cytiva Life Sciences) has a globular molecular weight range of 10 kDa to 600 kDa, making it ideal for separating larger biomolecules; however, separating 10 kDa from 20 kDa would not be ideal. Similarly, a Superdex 75 increase™ (Cytiva Life Sciences) can resolve 3 kDa to 70 kDa, making it ideal for the above situation. However, if a solution contained three macromolecules, 10 kDa, 65 kDa, and 250 kDa, it would be difficult to find a resin that could effectively separate the three macromolecules. Field-flow fractionation or asymmetric-flow field-flow fractionation (A4F) has recently emerged as a modern alternative to SEC. A4F combines multiple physical forces to separate molecules, whereas SEC relies exclusively on the hydrodynamic radius (R_h). A4F subjects a heterogenous solution of macromolecules to a combination of size and charge

exclusionary forces. 1) The first portion of the flow cell subjects the solution to cross flow, whereas a flow is perpendicular to the normal flow via a membrane which causes the molecules to concentrate on the membrane. 2) Diffusion then acts as a counterbalancing force that moves molecules away from the membrane proportionate to their translational diffusion coefficient (D_T) and R_h. 3) An anode at the top and a cathode at the bottom creates a charge difference, resulting in a change in retention time based on the zeta potential of molecules in the solution. 4) The final separating force is the basal laminar flow, which is slightly different based on the height of the particles in the chamber, resulting from the other three forces (14). This combination of multiple separation forces results in the ability of A4F to perform high-resolution separation of molecules between 1 and 1000 nm and beyond (15).

2.4 Application

MALS experiments are fundamental in determining sample quality and oligomeric properties. Often, electrophoresis gels are the gold standard regarding sample quality. SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) is widely used to assess the quality of recombinantly purified proteins but suffers from multiple problems. Firstly, the technique is denaturing, so it is impossible to evaluate the oligomeric interactions of the desired product. Secondly, all size determination is relative, based on a ladder of known protein sizes. Lastly, samples are often heated to 95°C, which could induce degradation in some proteins (16). These downsides make MALS, specifically SEC-MALS or A4F-MALS, a better alternative. MALS experiments are performed under solution conditions, so molecules can be examined in conditions as close to physiological as *in vitro* experiments allow. Allowing molar mass measurements in solution also allows for the distinctions between oligomeric species, something SDS-PAGE cannot do. Furthermore, MALS is the ideal tool to determine the molecular weight in solution; the measurement is absolute and based on the light scattering of the molecules, an intrinsic physical property. MALS has additionally found wide use in industry because of its ease of use, potential for high throughput, and reproducibility.

Modern chromatography systems often utilize 96 well plates or even 384 well plates, allowing for a larger number of samples to be run sequentially. Additionally, an individual can set up multiple SEC-MALS experiments back-to-back, often running a single SEC-MALS instrument for days, significantly reducing downtime. However, as mentioned previously, considerations of proper SEC columns need to be of fundamental concern meaning two molecules of similar sizes will likely be impossible to distinguish using SEC-MALS. A4F incorporates separation based on charge, which could theoretically separate two similarly sized molecules if their charge difference is substantial enough. In addition to A4F, lately, IEX-MALS (ion exchange coupled multi-angle light scattering) has been gaining popularity as an alternative to simply separating solutions based on size and shape (17). A wide variety of systems have utilized MALS in their characterization, with some recent examples being: The interaction of liposomes with bile salts to assess the stability of oral drug carriers (18), modeling long-term stability of monoclonal antibodies (19), characterizing adeno-associated virus vectors using IEX coupled MALS (20), the characterization of recombinantly purified inorganic pyrophosphatase for use in *in vitro* transcription (21), characterization of human DNA ligase IIIα complexes after expression in insect cells (22), characterization of SARS-COV-2 spike protein reference material (23), and to measure molecular weight, protein and polysaccharide concentrations in pneumococcal conjugate vaccine formulations (24).

2.5 Performing the MALS Experiments & Technical Considerations.

All technical and technique considerations are from my experience throughout the years working with this technique. Figure 2.2 is a picture of the system on which I performed all my MALS experiments. Figure 2.2A is the Thermo Scientific Vanquish UHPLC (ultra high-performance liquid chromatography), the chromatography pump used to inject samples and pump the liquid phase solvent throughout the system. The first consideration is that the light-scattering instrument (Figure 2.2C) needs to have an uninterrupted and consistent flow rate throughout the experiment. Initially, experiments were performed on an AKTA Pure® FPLC

(fast protein liquid chromatography) from Cytiva Life Science, which proved somewhat problematic. The primary problem was that the injection valve needed to switch between load and injection, effectively causing a quick change in pressure and an interruption to the buffer equilibration. The Vanquish UHPLC solved this problem through an injection arm, which injects the sample directly into a continuous flow, causing no interruption in pressure. These pressure issues are evident in Figure 2.3, where panel A shows the raw scattering intensity of both a 5 mg/mL and 12 mg/mL BSA (bovine serum albumin) sample injected into the AKTA Pure FPLC. Both injections show a significant scattering intensity spike at the beginning of the run. However, with the Vanquish, the ability to inject the sample into a continuous flow completely removes this spiking of intensity (Figure 2.3B).

The second challenge with the AKTA FPLC is the size of the typical SEC column and injection volume. Typically, an SEC column is ~24 mL, with an injection volume of 500 µL.

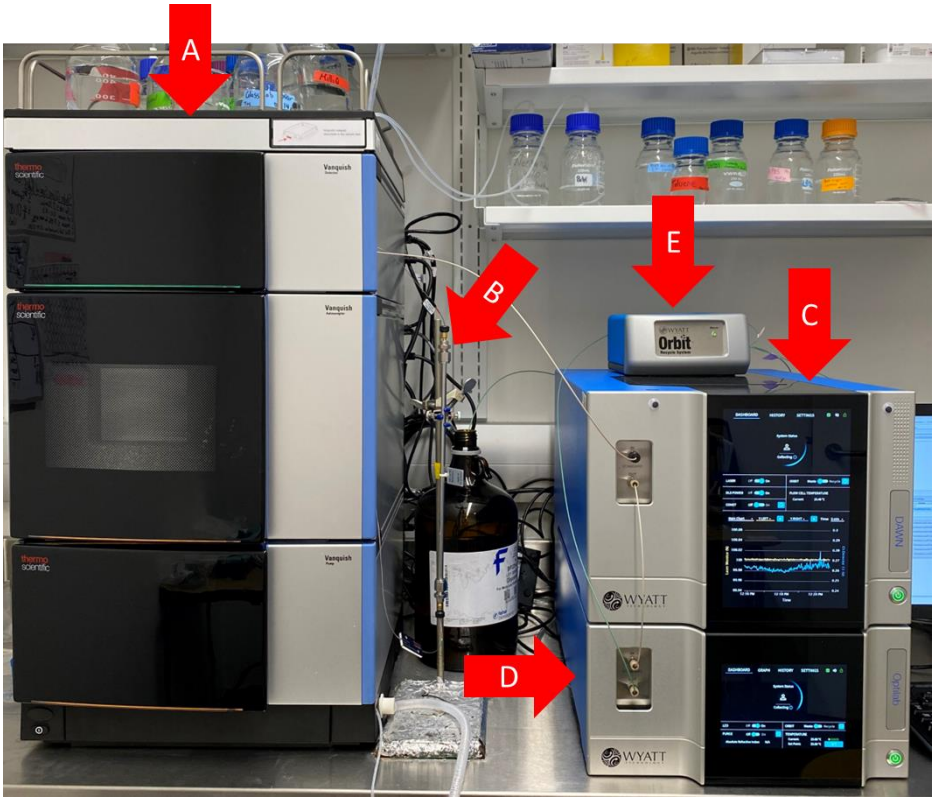


Figure 2.2 SEC-MALS setup in SA9156 of Dr. Trushar Patel's research lab. A) Thermo Scientific Vanquish UHPLC chromatography pump. **B)** Shodex 403-4F KW size exclusion chromatography column specific for HPLC systems. **C)** Wyatt Technology™ DAWN® multi angle light scattering instrument. **D)** Wyatt Technology Optilab® dRI detector. **E)** Wyatt Technology™ Orbit Recycle System.

MALS requires a decently high sample concentration to reduce the background noise, and injecting 500 µL into a 24 mL column results in a significant dilution of the original sample. Changing to the Vanquish UHPLC resulted in the ability to load only 100 µL of sample into an ~5.9 mL SEC column (Figure 2.2B). While the dilution is approximately the same, the total sample required is approximately 5x less. The second consideration is with the Optilab® (Figure 2.2D, which is the in-line refractometer). It is important to properly equilibrate the entire system to reduce the noise of the light-scattering signal and match the reference cell to the flow cell of the refractometer. The "purge" function on the refractometer modifies the flow path towards the reference cell and away from the flow cell. This purge, I would argue, is as crucial as equilibrating the flow path, and the time spent on this step should be equal to the time spent equilibrating the flow path. Figure 2.4 represents two separate experiments done on BSA, one with a short equilibration of the reference cell (panel A) and one with a long equilibration of the reference cell (panel B). Both graphs represent detector voltage ranges of 0.0002, but Figure 2.4B is a much cleaner representation of BSA than Figure 2.4A. Since molecular weight determination depends on concentration determination (a dRI (refractive index) measurement in this case), figure 2.4B results in a much more significant amplitude change from baseline than Figure 2.4A, resulting in a much more accurate

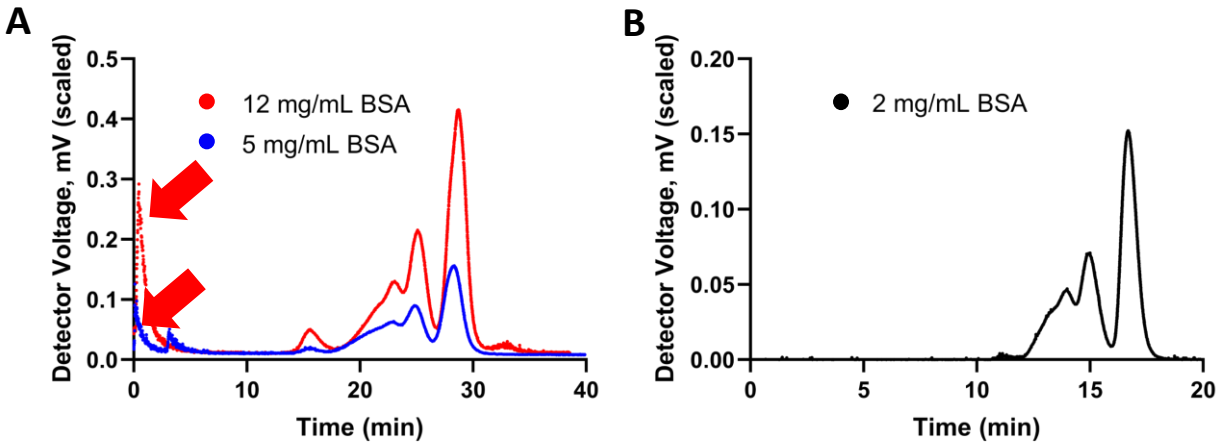


Figure 2.3. Comparison of AKTA Pure FPLC and Vanquish UHPLC SEC-MALS raw scattering data. **A)** Data collected on the AKTA Pure FPLC and a Superdex 200 increase column at 0.4 mL/min. Arrows represent an interruption in scattering intensity because of a change in pressure upon sample injection. **B)** Data collected on Vanquish UHPLC and a Shodex 403-4F KW SEC column at 0.160 ml/min.

molecular mass determination. Another consideration is the Orbit recycle system (Figure 2.2E); theoretically, this can save buffer by recycling the same buffer you use to equilibrate the system, essentially making a complete loop. I would only recommend using this if you have components that are very costly in your buffer. Running buffer consistently through meters of tubing is more likely to cause contamination. Unfortunately, in switching between buffers between experiments, it is difficult to know if the lines have been sufficiently purged not to introduce unwanted buffer components to your equilibration. Finally, my last technical consideration for SEC-MALS is always including a BSA standard before any experiment. Often, samples do not form perfectly homogenous peaks with easily interpreted known molar masses. Including a BSA standard provides valuable information to assess the quality of your equilibration, column, and overall experimental conditions. If the BSA standard data do not look as expected and do not return a monomeric molecular mass of 66 ± 5% kDa (25), then your experimental setup or instrument should be re-evaluated before forming experimental conclusions from any other samples run after that.

Ultimately, SEC-MALS is a highly efficient and valuable biophysical tool to evaluate sample quality, oligomerization state, and accurate molar mass. The absolute molar mass provided by the technique is independent of the size and shape of the molecule, so

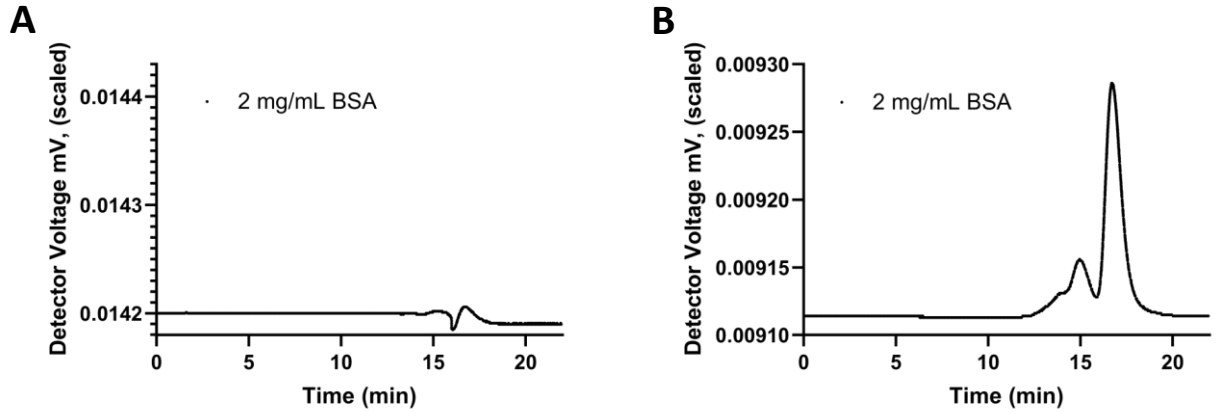


Figure 2.4. Refractive index measurements comparing equilibration of the reference cell. A) Refractive index measurement of 2 mg/mL BSA in 1x PBS with ~15 min reference cell equilibration. **B)** Refractive index measurement of 2 mg/mL BSA in 1x PBS with ~2h reference cell equilibration.

theoretically, all molecules are feasible candidates for SEC-MALS analysis, including globular proteins, fibrils, polysaccharides, carbohydrate chains, nucleic acids, and other non-biological molecules.

2.6 References

1. Zimm, B.H. (1948) The Scattering of Light and the Radial Distribution Function of High Polymer Solutions. *The Journal of Chemical Physics*, **16**, 1093-1099.
2. Einstein, A. (1910) Theorie der Opaleszenz von homogenen Flüssigkeiten und Flüssigkeitsgemischen in der Nähe des kritischen Zustandes. *Annalen der Physik*, **338**, 1275-1298.
3. Strutt, J.W. (1910) The incidence of light upon a transparent sphere of dimensions comparable with the wavelength. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, **84**, 25-46.
4. Debye, P. (1944) Light Scattering in Solutions. *Journal of Applied Physics*, **15**, 338-342.
5. Doty, P.M., Zimm, B.H. and Mark, H. (1945) An Investigation of the Determination of Molecular Weights of High Polymers by Light Scattering. *The Journal of Chemical Physics*, **13**, 159-166.
6. Zimm, B.H. (1948) Apparatus and Methods for Measurement and Interpretation of the Angular Variation of Light Scattering; Preliminary Results on Polystyrene Solutions. *The Journal of Chemical Physics*, **16**, 1099-1116.
7. Fixman, M. (1955) Excluded Volume in Polymer Chains. *The Journal of Chemical Physics*, **23**, 1656-1659.
8. Pam Wang, R.A., Michelle Chen, Kristine Legaspi. (2020), Wyatt Technologies, pp. 1-4.
9. Zhao, H., Brown, P.H. and Schuck, P. (2011) On the distribution of protein refractive index increments. *Biophysical journal*, **100**, 2309-2317.
10. Swinehart, D.F. (1962) The Beer-Lambert Law. *Journal of Chemical Education*, **39**, 333.
11. Sobral, H. and Peña-Gomar, M. (2015) Determination of the refractive index of glucose-ethanol-water mixtures using spectroscopic refractometry near the critical angle. *Appl. Opt.*, **54**, 8453-8458.
12. Glover, F.A. and Goulden, J.D.S. (1963) Relationship between Refractive Index and Concentration of Solutions. *Nature*, **200**, 1165-1166.
13. Wyatt, P.J. (1997) Multi-angle Light Scattering: The Basic Tool for Macromolecular Characterization. *Instrumentation Science & Technology*, **25**, 1-18.
14. Plavchak, C.L., Smith, W.C., Bria, C.R.M. and Williams, S.K.R. (2021) New Advances and Applications in Field-Flow Fractionation. *Annual Review of Analytical Chemistry*, **14**, 257-279.
15. Johann, C. and Deng, C. (2016) Innovations in FFF – Eclipse NEON.
16. Bikaki, M., Shah, R., Müller, A. and Kuhnert, N. (2021) Heat induced hydrolytic cleavage of the peptide bond in dietary peptides and proteins in food processing. *Food Chemistry*, **357**, 129621.
17. Amartely, H., Avraham, O., Friedler, A., Livnah, O. and Lebendiker, M. (2018) Coupling Multi Angle Light Scattering to Ion Exchange chromatography (IEX-MALS) for protein characterization. *Scientific reports*, **8**, 6907.
18. Bohsen, M.S., Tychsen, S.T., Kadhim, A.A.H., Grohganz, H., Treusch, A.H. and Brandl, M. (2023) Interaction of liposomes with bile salts investigated by asymmetric flow field-flow fractionation (AF4): a novel approach for stability assessment of oral drug carriers. *Eur J Pharm Sci*, 106384.
19. Legrand, P., Dufaÿ, S., Mignet, N., Houzé, P. and Gahoual, R. (2023) Modeling study of long-term stability of the monoclonal antibody infliximab and biosimilars using liquid-chromatography-tandem mass spectrometry and size-exclusion chromatography-multi-angle light scattering. *Anal Bioanal Chem*, **415**, 179-192.
20. Wagner, C., Innthaler, B., Lemmerer, M., Pletzenauer, R. and Birner-Gruenberger, R. (2022) Biophysical Characterization of Adeno-Associated Virus Vectors Using Ion-Exchange Chromatography Coupled to Light Scattering Detectors. *Int J Mol Sci*, **23**.

21. Tersteeg, S., Mrozowich, T., Henrickson, A., Demeler, B. and Patel, T.R. (2022) Purification and characterization of inorganic pyrophosphatase for in vitro RNA transcription. *Biochem Cell Biol*, **100**, 425-436.
22. Rashid, I., Tsai, M.S., Sverzhinsky, A., Hlaing, A.S., Shih, B., Thwin, A.C., Lin, J.G., Maw, S.S., Pascal, J.M. and Tomkinson, A.E. (2022) Purification and Characterization of Human DNA Ligase IIIα Complexes After Expression in Insect Cells. *Methods in molecular biology (Clifton, N.J.)*, **2444**, 243-269.
23. Stocks, B.B., Thibeault, M.P., Schrag, J.D. and Melanson, J.E. (2022) Characterization of a SARS-CoV-2 spike protein reference material. *Anal Bioanal Chem*, **414**, 3561-3569.
24. Deng, J.Z., Lancaster, C., Winters, M.A., Phillips, K.M., Zhuang, P. and Ha, S. (2022) Multi-attribute characterization of pneumococcal conjugate vaccine by Size-exclusion chromatography coupled with UV-MALS-RI detections. *Vaccine*, **40**, 1464-1471.
25. (2016) In Aronson, J. K. (ed.), *Meyler's Side Effects of Drugs (Sixteenth Edition)*. Elsevier, Oxford, pp. 1045.

Chapter 3. Microscale Thermophoresis

3.1 Foreword

Chapters 2-5 are written because of my extensive focus on learning and applying numerous biophysical techniques throughout my Ph.D. While this is not an extensive list, I primarily focused on four techniques, MST, SEC-MALS, SAXS, and AUC. This chapter focuses on MST and contains a manuscript titled "Microscale Thermophoresis: Warming up to a New Biomolecular Interaction Technique" I wrote in collaboration with Dr. Vanessa Meier-Stephenson and Dr. Trushar R Patel, published in *The Biochemist* Volume 41, Issue 2, Pages 8-12 on April 1, 2019. https://doi.org/10.1042/BIO04102008. This paper describes how MST works, the information obtained from MST experiments, its advantages and disadvantages, and the diagnosis of common problems. This publication is reproduced with permission from Portland Press Ltd and the Creative Commons CC by license and re-formatted to fit thesis formatting.

3.2. Author List and Affiliations

Tyler Mrozowich[1], Vanessa Meier-Stephenson[1], and Trushar R Patel[1]

1. Alberta RNA Research and Training Institute, Department of Chemistry and Biochemistry, University of Lethbridge, 4401 University Drive, Lethbridge, Alberta T1K 3M4, Canada)

3.3 Abstract

Biomolecules, such as RNA, DNA, proteins and polysaccharides are at the heart of fundamental cellular processes. These molecules differ greatly with each other in terms of their structures and functions. However, in the midst of the diversity of biomolecules is the unifying feature that they interact with each other to execute a viable, biological system. Interactions of biomolecules are critical for cells to survive and replicate, for food metabolism to produce energy, for antibiotics and vaccines to function, for spreading of diseases, and for every other biological process. An improved understanding of these interactions is crucial for studying how cells and organs function, to appreciate how diseases are caused, how infections

occur, with infinite implications in medicine and therapy. Many biochemical and biophysical techniques are currently being employed to study biomolecular interaction. Microscale Thermophoresis is a relatively new biophysical technique that can provide powerful insight into the interactions of biomolecules and is quickly being adopted by an increasing number of researchers worldwide. This article provides a brief description of principles underpinning MST process, benefits and limitations.

3.4 What is MST? And how does it work?

Microscale thermophoresis is a biophysical technique that measures how strongly biomolecules interact based on their degree of dispersion upon heating ("thermo" = heat; "phoresis" = migration). Essentially, one of the biomolecules (a protein, for example) is labeled with a fluorophore designated the "target", while the other molecule (RNA, for example) is unlabelled and is mixed with the target in varying concentrations acting as the "ligand" (Figure 3.1). Next, the baseline fluorescence is measured, followed by a short

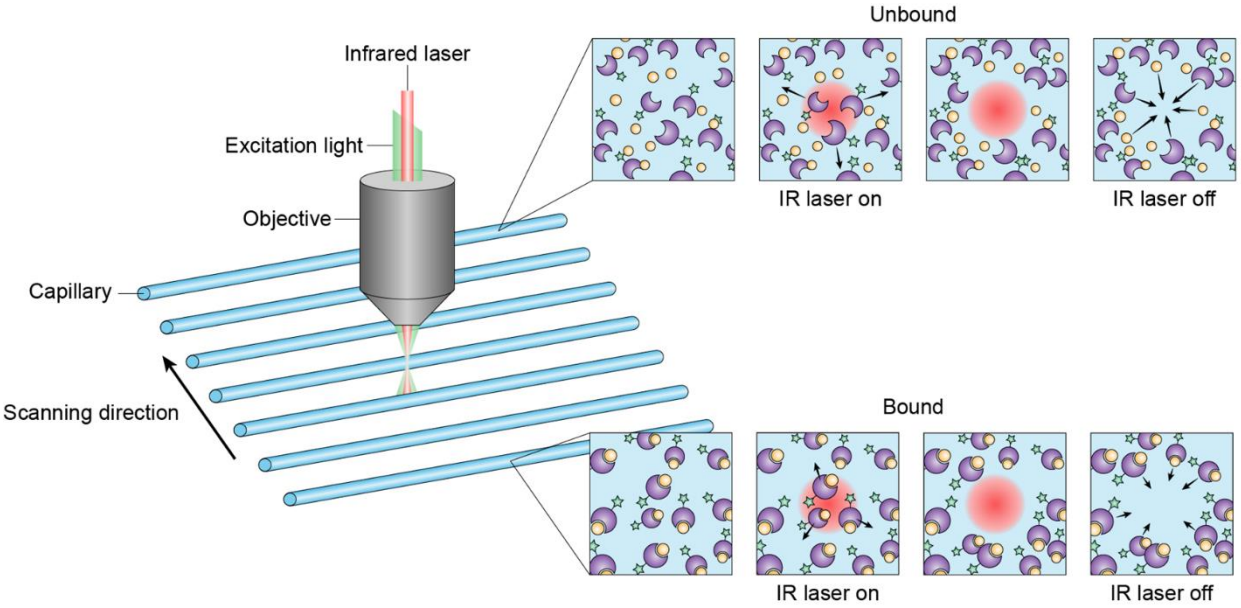


Figure 3.1. Microscale thermophoresis mechanism of action. Diagram of the general mechanism of MST. An IR laser is used to heat the sample which will either contain molecules in a bound or unbound state. Upon stimulation with the IR laser (shown in the "IR laser on" panels), the molecules will diffuse away. Once the IR laser is turned off, the molecules will diffuse back to the IR stimulated area (shown in the IR laser off panels). If the biomolecules bind, there will be a change in the diffusion pattern once the IR laser is turned on, compared with the unbound state. The biomolecular diffusion is measured as a change in fluorescence output of a fluorophore (represented by the green star) attached to the target molecule

exposure with an infrared laser (heat source) on each sample; the change in fluorescence between the "hot" and "cold" state of each sample is detected. If the ligand is interacting with the fluorescently-labeled target molecule, the complexes will diffuse at a different rate when heated up relative to the individual components alone. These micro-movements (Figure 3.1) are detected by the machine as a drop-in fluorescence and are plotted as such.

3.5 What kind of information can we obtain from MST?

MST has a variety of applications. Focusing on the features of greatest interest to the biochemist, MST can be used to determine dissociation constants (K_d), an equilibrium measurement of the likelihood of a complex to dissociate at a certain concentration, in reference to interacting biomolecules. This can be useful to study affinity between different

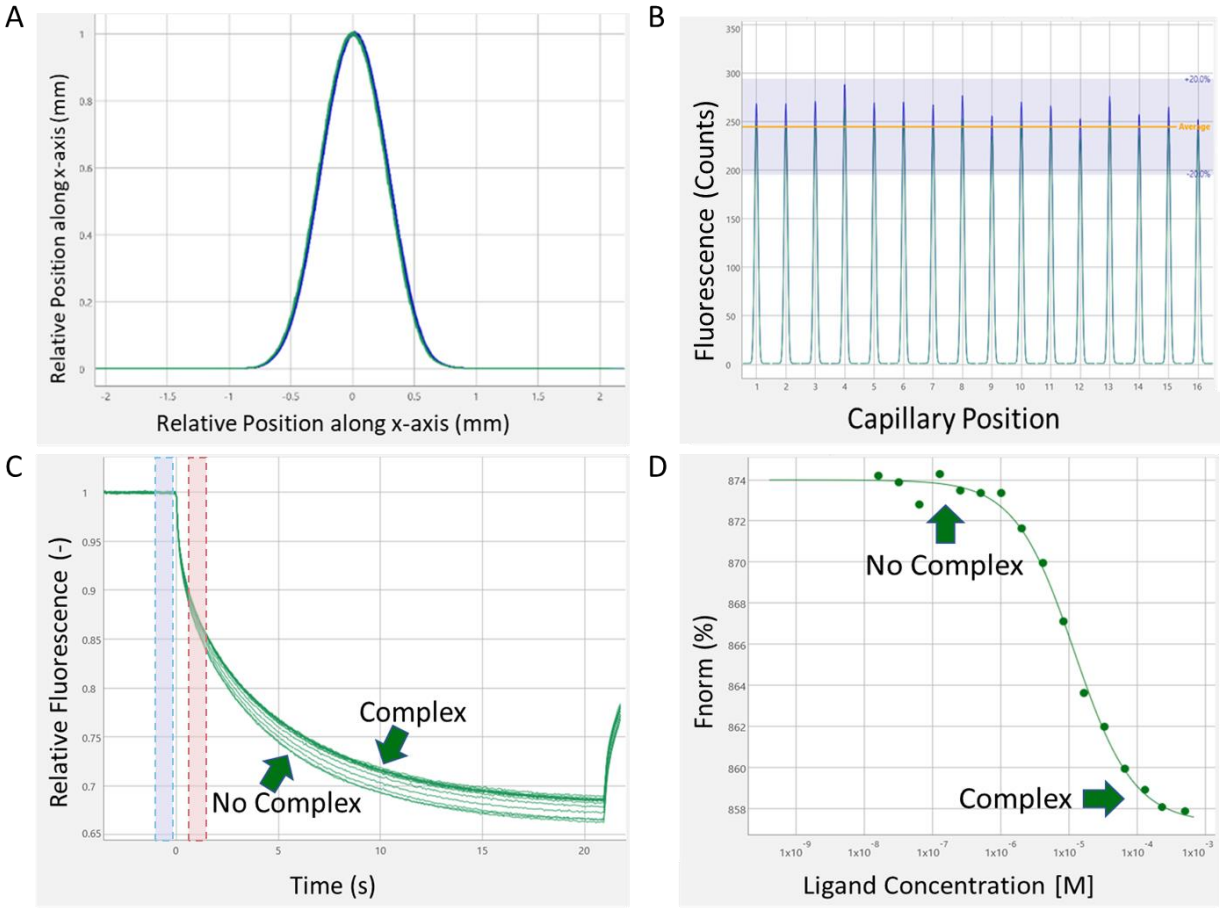


Figure 3.2 Microscale thermophoresis outputs. Plots visualizing several MST outputs from different experiments. **A)** Pretest capillary scan comparison. **B)** Capillary scan during binding affinity test. **C)** MST trace output during binding affinity test. **D)** Binding curve output after binding affinity test.

systems such as protein-protein (e.g., antigen-antibody interactions), RNA-protein complexes (e.g. ribosomal studies), and small-molecule-protein (e.g., drug binding studies). For example, using MST, researchers probed an interaction between a small molecule drug (cetylpyridinium chloride) with hepatitis B viral dimeric nucleocapsid protein HBcAg to use this novel drug as a potential inhibitor for viral replication. MST can also probe nucleic acid-protein interactions. For example, researchers investigated the binding of G-quadruplex RNA aptamer AIR-3 with the human interleukin-6 receptor that plays a role in cancer and inflammatory diseases. With the assistance of MST, the researchers aimed to understand the RNA structure, shape, and RNA interaction site and binding stoichiometry. Enzyme kinetics

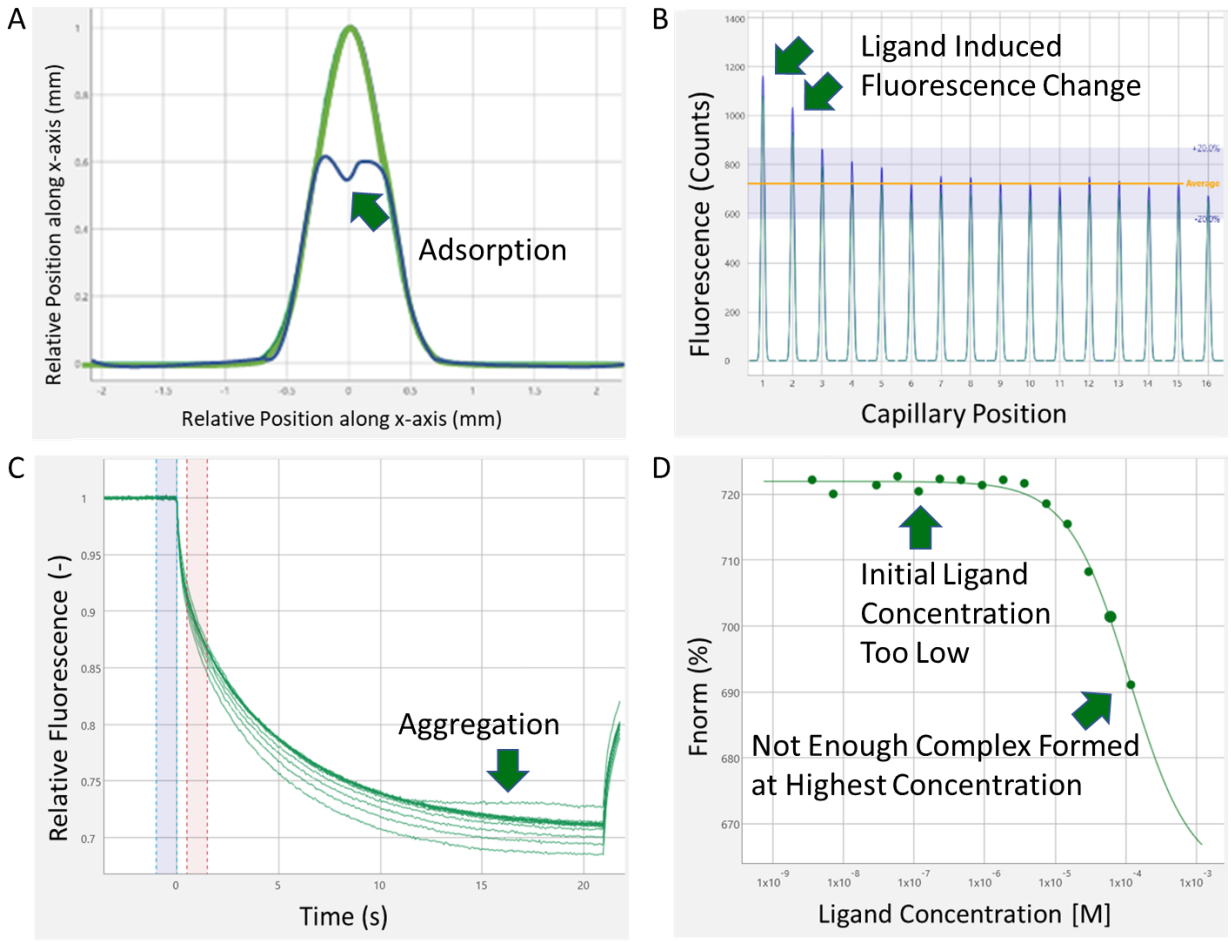


Figure 3.3. Visualizing common problems. Plots representing some common errors that can arise during a complete MST experiment. **A)** Adsorption during a pretest experiment. **B)** Fluorescence inhomogeneity during a binding affinity test. **C)** Aggregation of sample during a binding affinity test. **D)** Incomplete binding curve after a binding affinity test.

and competition studies can also be conducted. Importantly, MST experiments are carried out in solution and are able to mimic biologically relevant environments. Buffer choice is essentially unlimited (as long as the target and ligand molecules are 'happy' with the buffer conditions) and samples could even be run in serum or cell lysates to detect a biomolecular interaction, which can provide unique information, only otherwise available via *in vivo* assays.

3.6 What are the advantages and disadvantages of MST?

One of the important advantages of MST over other techniques is the range of biomolecules that can be studied: from sugars or small molecule drugs up to large proteins, enzymes and even more complex structures such as ribosomes. The versatility of samples that can be studied makes this a highly attractive technique in the field. MST also has advantages over other fluorescence-based techniques, including the simplicity of measurement, the low sample volume required, and the infinite variety of buffer formulations supported. The technique can also be used to provide insight into hydration shell dynamics, molecular charge interactions and changes in size of biomolecules during interaction. The subtleties of such specific applications will not be described here, but those interested can refer to the additional readings below. As well, should fluorescent-labeling not be optimal or desired for the system, there are adaptations that can be made for non-fluorescent approaches.

As with any technique, there are negative aspects that users should consider. The initial cost of the machine and upfront expense of dyes can be a hurdle, however, given the smaller volume of sample required, some of these costs can be offset by the reduced amount of high-quality starting sample needed. Fluorescent labeling, in general, can be a negative aspect because of the potentially time-consuming process of efficiently labeling your target. This may be partially circumvented by using fluorophores previously shown to provide reliable results (See Fluorophores below). MST does not produce detailed information about the shape or location of the binding. To do this, one would need to consider incorporating complementary

studies such as X-ray crystallography, nuclear magnetic resonance or small-angle X-ray scattering.

3.7 Pioneer Voyage

3.7.1 Sample quality

As with other biophysical techniques, one of the most critical things is to start with high quality, monodispersed samples (i.e. with particles of uniform size). Often, this requires multiple different methods such as; gel electrophoresis, static light scattering, and more recently, using the NanotemperTM Tycho NT.6 which employs differential scanning fluorimetry to measure protein foldedness.

3.7.2 Fluorophore

After sample quality is confirmed, the next step is to determine which fluorophore will best suit the experiment. MST utilizes differences in thermal shift by detecting fluorescence to evaluate biomolecular interaction, therefore only one of your interacting molecules needs to be fluorescently labeled. For nucleic acids, Cy5 fluorescent labeling is reliably detected by the MST sensor. Companies that synthesize nucleic acid sequences will provide Cy5 labeling as an option, and thus would be MST-ready. For proteins, using the RED-tris-NTA red fluorescent dye sold by Nanotemper™ is a simple way to label Histidine-tag reconstituted proteins. Dyes that can efficiently label primary amines like those found in lysines or dyes specific to labeling of cysteine residues are also options. It is also extremely important to thoroughly evaluate the fluorescent labeling procedure used, to ensure that your target molecule is sufficiently labeled. Unlabeled target will still interact with the ligand but will not be measured by the MST instrument resulting in an inaccurate K_d value. Determining the labeling efficiency will also help verify whether the experimental concentrations will be sufficiently above the dissociation constant so the target will stay fluorescently labeled.

3.7.3 Buffer

An important consideration when choosing a biophysical technique is the buffer conditions in which the experiment takes place in. MST is extremely versatile, supporting almost any buffer condition. There are, however, a few modifications to the experimental buffer that may improve the MST results. The first is adding Tween-20 (0.05% final concentration) to help mitigate protein adsorption and aggregation. The second is using bovine serum albumin (BSA; 0.4 mg/mL) in the buffer to prevent protein adsorption to microcentrifuge tubes, where the BSA will coat the tube, allowing free diffusion of the target and ligand molecules.

3.7.4 Starting concentrations

Following labeling and buffer selection, the next step is to determine the concentrations of the target and ligand to be used. The target concentration must be high enough with the attached fluorophore to be detectable by the MST machine. This is performed in the "pretest" option of the software, in which the machine will scan two capillary tubes containing only target solution, to determine if the fluorescence is sufficient to continue (Figure 3.2A). Performing the pretest is highly recommended, as identification of insufficient fluorescence can alert the user to make adjustments to the labeling or concentrations, minimizing wasted materials from a failed or indeterminant binding affinity experiment. After the pretest, if binding has been previously shown (i.e., via an alternate biophysical method), one can skip the binding check step. If binding is not known, it is advisable to perform the software's binding check. This will provide a 'yes-no' answer as to whether there is detectable binding and requires less sample than a full binding affinity. Binding checks require four sample capillaries of target only, and four sample capillaries of target mixed with a high concentration of ligand. After all the above parameters have been investigated, the binding affinity test can be done.

3.7.5 Binding affinity measurement

To get an accurate binding curve, it is advisable to begin with an initial ligand concentration that is 20x the K_d. However, often the K_d will be unknown, therefore it is generally a good idea to start with a ligand concentration that is several-fold higher than the target (20-30 times) so as to observe the full range of the interaction. A 16-tube serial dilution of ligand is performed, and then a fixed amount of target is added to each tube. Pipetting accuracy in the previous steps is generally not an issue, but to gain an accurate binding curve one must be fully confident in the volumes pipetted. This can be visualized by reviewing the capillary scans output provided by the machine and ensuring there are no fluctuation (Figure 3.2B). Although only 10 µL of sample is required per capillary, as a pipetting error can be more prominent with the smaller volumes used, we recommend mixing samples to final volumes of 20uL using low-bind pipette tips. A raw MST trace plot should ideally show a change in relative fluorescence as ligand concentration increases (Figure 3.2C). An accurate binding curve should have a few points at high concentration representing complete binding; a few points at concentrations with no binding; and ideally, the remaining points defining the slope of the binding curve (Figure 3.2D).

3.7.6 Diagnosis of Common Errors.

It is important to be able to understand what potential problems will present visually in MST results. Target adsorption can present itself similar as a warped Gaussian peak in initial pretest scans (Figure 3A). As mentioned above, BSA addition to the buffer is often ideal in preventing this. Fluorescence inhomogeneity in capillary scans is harder to diagnose (Figure 3.3B), but the most common issue is improper fluorescence binding to the target. This is especially common when EDTA (ethylenediaminetetraacetic acid) is present in the buffer, and a Histidine-tag binding dye or other metal-ion dependent marker is being used. Changing the buffer to one without EDTA can help mitigate this. Aggregation is easily diagnosed (Figure 3.3C), but less easy to remedy. If possible, use a lower concentration of both ligand and target, and add detergent (e.g. Tween-20). Incomplete binding curves are identified easily,

and the solution is to use higher (or lower) concentrations of the ligand (Figure 3.3D). If the problem persists after increasing ligand concentration, the efficiency of the labeling of fluorescent dye to the target should be re-evaluated.

3.8 Concluding Remarks

While learning a new technique such an MST may come across as a challenging task, once the fundamentals are understood and the potential pitfalls are properly diagnosed, MST can be a powerful biophysical technique that an increasing number of researchers are turning towards. MST has proven to be an efficient, cost-effective and reliable technique which has the potential for continuous improvement as optics and fluorescence continue to progress. An exciting future for MST is the continuing research into label-free MST experiments which can utilize intrinsic fluorescence, such as that generated by aromatic amino acids. Given the astounding amount of biomolecular interactions that a single cell performs every second, researchers need every tool possible to dissect the infinite complexity that nature has evolved to generate and sustain life. MST will continue to be an invaluable asset in the continuing search for this knowledge.

3.9 Further Reading

1. Seo, H.W., et al., Cetylpyridinium chloride interaction with the hepatitis B virus core protein inhibits capsid assembly. Virus Research, 2019. 263: p. 102-111.
2. Szameit, K., et al., Structure and target interaction of a G-quadruplex RNA-aptamer. RNA biology, 2016. 13(10): p. 973-987.
3. Ascher, D.B., et al., Potent hepatitis C inhibitors bind directly to NS5A and reduce its affinity for RNA. Sci Rep, 2014. 4: p. 4765.
4. Asmari, M., et al., Thermophoresis for characterizing biomolecular interaction. Methods, 2018. 146: p. 107-119.
5. Dijkman, P.M. and A. Watts, Lipid modulation of early G protein-coupled receptor signalling events. Biochimica et Biophysica Acta (BBA) - Biomembranes, 2015. 1848(11, Part A): p. 2889-2897.
6. Moon, M.H., et al., Measuring RNA-Ligand Interactions with Microscale Thermophoresis. Biochemistry, 2018. 57(31): p. 4638-4643.
7. 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.
8. Rodrigo-Unzueta, A., et al., Molecular Basis of Membrane Association by the Phosphatidylinositol Mannosyltransferase PimA Enzyme from Mycobacteria. Journal of Biological Chemistry, 2016. 291(27): p. 13955-13963.
9. Sparks, R.P. and R. Fratti, Use of Microscale Thermophoresis (MST) to Measure Binding Affinities of Components of the Fusion Machinery. Methods Mol Biol, 2019. 1860: p. 191-198.
10. Wienken, C.J., et al., Protein-binding assays in biological liquids using microscale thermophoresis. Nat Commun, 2010. 1: p. 100.
11. Seidel, S.A., et al., Label-free microscale thermophoresis discriminates sites and affinity of protein-ligand binding. Angew Chem Int Ed Engl, 2012. 51(42): p. 10656-9.
12. Patel, T.R., et al., Structural studies of RNA-protein complexes: A hybrid approach involving hydrodynamics, scattering, and computational methods. Methods, 2017. 118-119: p. 146-162.

Chapter 4. Small Angle X-Ray Scattering

4.1 Foreword

Chapters 2-5 are written because of my extensive focus on learning and applying numerous biophysical techniques throughout my Ph.D. While this is not an extensive list, I primarily focused on four techniques, MST, SEC-MALS, SAXS, and AUC. This chapter focuses on SAXS and contains a manuscript titled "Structural Studies of Macromolecules in Solution using Small Angle X-Ray Scattering" I wrote in collaboration with Steffane McLennan, Dr. Michael Overduin and Dr. Trushar R Patel, published in *The Journal of Visualized Experiments,* Volume 141, on November 5, 2018. https://dx.doi.org/10.3791/58538. This publication is unique because it is accompanied by a descriptive video (linked at the end of this chapter). The manuscript describes the basics of SAXS, and a detailed description of data analysis following data collection, as well as experimental results for nidogen-1 and laminin γ-1. This publication is reproduced with permission from MyJoVE and the Creative Commons CC by license and re-formatted to fit thesis formatting.

4.2 Author list and Affiliations

Tyler Mrozowich[1], Steffane McLennan[2], Michael Overduin[2, 4], and Trushar Patel[1, 3, 4]

1. Alberta RNA Research and Training Institute, Department of Chemistry and Biochemistry, University of Lethbridge, 4401 University Drive, Lethbridge, Alberta T1K 3M4, Canada
2. Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7, Canada
3. Department of Microbiology, Immunology and Infectious Diseases, Cumming School of Medicine, University of Calgary, 2500 University Dr. NW Calgary, Alberta, T2N 1N4, Canada
4. DiscoveryLab, Faculty of Medicine & Dentistry, University of Alberta, Edmonton, Alberta, T6G 2H7, Canada

4.3 Abstract

Protein-protein interactions involving proteins with multiple globular domains present technical challenges for determining how such complexes form and how the domains are

oriented/positioned. Here, a protocol with the potential for elucidating which specific domains mediate interactions in multicomponent system through *ab initio* modeling is described. A method for calculating solution structures of macromolecules and their assemblies is provided which involves integrating data from small angle X-ray scattering, chromatography, and atomic resolution structures together in a hybrid approach. A specific example is that of the complex of full-length nidogen-1 which assembles extracellular matrix proteins and forms an extended, curved nanostructure. One of its globular domains attaches to laminin γ-1 which structures the basement membrane. This provides a basis for determining accurate structures of flexible multidomain protein complexes and is enabled by synchrotron sources coupled with automation robotics and size exclusion chromatography systems. This combination allows rapid analysis in which multiple oligomeric states are separated just prior to SAXS data collection. The analysis yields information on the radius of gyration, particle dimension, molecular shape and interdomain pairing. The protocol for generating 3D models of complexes by fitting high-resolution structures of the component proteins is also given.

4.4. Introduction

Cells contain intricate networks of proteins which act as molecular machines to carry out cellular functions such as signaling cascades and maintaining structural integrity. The ways in which these different components move and interact in three-dimensional space gives rise to the specific functions of the macromolecules. The importance of the protein structure, dynamics, and interactions in determining function has provided the need for continually evolving, complex techniques to measure these properties. Of these, nuclear magnetic resonance (NMR), X-ray crystallography (XRC) and more recently, cryo-electron microscopy (CEM) provide high-resolution structural information. However, XRC and CEM yield structures of one of many biomolecular states and lack information about the dynamics of the protein structure, while 3D structure determination by NMR is typically limited to smaller globular proteins. One way to overcome these limitations is to utilize small angle X-ray scattering to

generate molecular envelopes of large, multidomain or complexed systems, and combine the high-resolution rigid macromolecular structures to elucidate the global architecture and dynamic features.

SAXS produces low-resolution envelopes of macromolecular complexes with a resolution of approximately 10-20 Å (1), giving insight not only into the structure but also the dynamic characteristics that the complex displays. Although SAXS utilizes X-rays to uncover molecular structure, it is unlike XRC in that the random isotropic orientation of the particles in solution does not lead to diffraction, but rather to scattering, which cannot yield atomic resolution. Instead, an electron "envelope" of the macromolecule is generated that represents an average of the conformations that the macromolecule displays. This information can be used in direct fitting of previously solved atomic resolution structures to infer regions of flexibility in a single protein or subunit organization, or dynamics in a larger, multi-protein complex. SAXS data is collected at synchrotrons using high-energy monochromatic X-rays or from in-house sources which offer a weaker X-rays source requiring hours rather than seconds of sample exposure time (Figure 4.1). SAXS data is often collected from several samples with a single experimental setup and buffer, requiring an extended time to collect a round of useful data on a system. Samples should, therefore, be stable and non-aggregating for at least a few hours based on verifiable quality control methods such as dynamic light scattering (DLS) and/or analytical ultracentrifuge analysis to obtain high-quality SAXS data (2,3). Here we provide a practical description of SAXS, the principles behind its usage, benefits, limitations and sample preparation and focus heavily on data collection and analysis, along with touching briefly on *ab initio* modeling using the extracellular matrix proteins nidogen-1 and laminin γ-1 as an experimental example.

4.5. Principles, Benefits, and Limitations of SAXS

The guiding principles behind SAXS are relatively simple: a solution of the monodispersed preparation of macromolecule(s) of interest is placed within a capillary and is

exposed to a high energy monochromatic X-ray beam. The photons cause electrons of the atomic shell to begin oscillating resulting in a spherical wave being emitted of the same energy and wavelength. Since every electron will oscillate, a constant background will be achieved, and the resulting electron density of the macromolecule is contrasted to the background. The resulting scattering intensity is collected as a function of the scattering angle, 2Θ (Figure 1).

While other techniques such as XRC, NMR, and CEM provide structural information at the atomic level, there are multiple benefits to SAXS which other techniques cannot provide. SAXS can be performed in almost any buffer and does not require any special sample preparation. This is particularly important in studying the behavior and structure of macromolecules under varying conditions such as the presence or absence of mono- or divalent cations, or changes in pH (4,5). SAXS has the ability to provide information about flexible regions of a macromolecule (6), something the other listed techniques can struggle with. Therefore, SAXS can be used as a strong complimentary technique with the stable portions of a macromolecule being studied with XRC, NRM or CEM, and the entire macromolecule or complex analyzed in low resolution with SAXS and combined using various analysis tools such as FoXSDock (7) or CRYSOL (8). Since SAXS is a solution technique, it is often used to confirm if static structures such as those obtained from XRC are consistent in solution (6). SAXS also has the advantage of being a technique that requires a relatively small amount of sample investment (typically 50-100 µL) and a relatively small amount of experiment time (30 min-1 h).

The largest limitation of SAXS is the vulnerability to sample aggregation and/or degradation which can lead to incorrect structural predictions. An aggregation, even as low as 5%, can scatter light in very high amounts, leading to an overestimation of the maximal particle dimension (D_{max}) and radius of gyration (R_g). On the other hand, sample degradation can lead to an underestimate of molecular properties. This vulnerability arises from SAXS being an averaging technique which means that sample homogeneity is critical to achieving

reliable and reproducible results. Any sample that is to be analyzed by SAXS should, therefore, undergo multiple methods of purification and homogeneity checks such as: denaturing and native gel electrophoresis, size exclusion chromatography, dynamic light scattering and analytical ultracentrifugation. Often SAXS beamlines will run samples through high-performance liquid chromatography as a final quality control step before SAXS (SEC-SAXS) (3,9). SAXS data should be collected at multiple concentrations and the R_g of each data set should be compared, ensuring a close similarity to avoid interparticle interactions and aggregation, which results in an overestimation of particle dimensions, leading to inaccurate data analysis and modeling. Since scattering depends on both concentration and size, smaller macromolecules may require a more specific optimization of the concentration range. This is due to the Reciprocity Theorem where large sizes scatter towards small angles and small sizes towards large angles. This manifests in data collection where I_O is proportional to R^6 where R is the particle radius. A final limitation of SAXS is the potential for radiation damage to the sample during exposure which can lead to distortion of the data. It is good practice to compare sample quality before and after SAXS sample exposure to ensure this is not occurring.

4.6. SAXS Sample Preparation and Data Acquisition

SAXS experiments require homogeneous, stable and non-aggregating protein samples. Stability and oligomeric state can be observed with either size exclusion chromatography (SEC), DLS and/or AUC prior to data collection. As previously described, the nidogen-1 and laminin γ-1 samples were subjected to DLS analysis and tricine SDS-PAGE to visualize sample purity (10). Samples cover a range of concentrations (1-4 mg/mL) depending on their size, their solution behavior such as self-association and aggregation along with stability. Five concentrations of nidogen-1, (139 kDa), three of laminin γ-1 (109 kDa) and four of the SEC-purified equimolar complex were prepared as previously described (10).

SAXS data for was collected using an in-house Rigaku system that contains a 3-pinhole camera equipped with + 002 microfocus sealed tube (Cu Kα radiation at 1.54 Å) and Confocal

Max-Flux (CMF) optics operating at 40 W. The system is also equipped with a 200 nm multi-wire 2D detector for data collection. However, with the availability of modern synchrotrons in France, Germany, UK, USA, and other countries which provide access to an SEC-SAXS set-up which facilitates separation of a monodispersed preparation from possible aggregation/degradation, we now routinely collect data at synchrotron facilities. A recently published article on a DNA G-quadruplex (11) is an example of an SEC-SAXS data collection strategy. In this case, the SAXS data were collected in the range of $0.08 \leq q \leq 0.26$ Å for 3 hours for nidogen-1 (2.0, 2.5, 3.0, 3.5 and 4.0 mg/mL); the laminin γ-1 (1.5, 2.0 and 2.5 mg/mL) and their complex (0.8, 1.0, 1.25 and 1.5 mg/mL). Data for buffer and samples were reduced by means of processing software specific to the system prior to subtraction of the buffer contribution from protein data with the program PRIMUS/qt (12) (Figure 4.2A).

4.7 Protocol

4.7.1. Data Analysis

Currently, there are a few software packages that are useful for SAXS data analysis: ScÅtter (13) (download available at www.bioisis.net), bioXtas RAW (14) and the ATSAS suite (12). This section provides an overview of general steps to be taken when analyzing raw SAXS data using the ATSAS program suite and specific steps are taken from Mac OS ATSAS 2.8.1 download. Other programs can be used and are briefly discussed later.

4.7.2 Buffer Subtraction

Note: These steps are relevant for static SAXS samples only.

Select the "TOOLS" menu option in PRIMUS/qt and select the data files of interest (up to 13 at once) using the "SELECT" button. Be aware that data files must be in ASCII format in which the first column is the s-vector axis and the second column is the intensity. Repeat this step for data collected for the buffer itself by inserting this data into a second "TOOLS" menu.

Select "SUBTRACT" in the Data Processing window which will generate a subtracted scattering curve representing only scattering from the macromolecule of interest. Repeat this step for each concentration.

4.7.3 Guinier Analysis

To perform Guinier analysis, load a buffer subtracted scatter curve into PRIMUS/qt which is performed in the same way curves were loaded previously.

Click "GUINIER" which will proceed in opening the Primus Guinier Wizard: a plot of ln(I) vs q^2 will be displayed.

To obtain a preliminary R_g use the "AUTORG" function, which is an external module built into PRIMUS/qt. Find the command prompt window and enter "autorg [datafile]".

Input multiple files at once by entering the same command prompt, but putting a space between each datafile name or by inputting all files in the selected folder by typing "autorg *.dat" which inputs every file ending in .dat.

Use the Guinier plot created earlier to assess the data quality; the green line under the Guinier plot shows the residuals plot representing a linearity of the fit. Be aware that non-linearity in the Guinier analysis could be a sign of sample aggregation and further analysis should not be performed in this case. (Note: A linear Guinier fit gives an R_g with a small error (<5%) and suggests a high-quality sample.)

4.7.4 Kratky Analysis

Load the data to be visualized in a similar manner as described above.

Click on the "SELECT" next to the data file name followed by "PLOT". This will plot the data in a separate window.

Click the "SASPLOT" button below the "PLOT" button.

Click "VIEW" and subsequently select "Y*s^2::X" which will plot the data as "q^2 x I(q) vs q". Be aware that globular proteins display a Gaussian peak while unfolded proteins will display a plateau instead of a peak and resemble a hyperbolic plot (15).

4.7.5 Data Merging

Load buffer subtracted data for each concentration in PRIMUS/qt once again like what has been previously described.

To merge the data, simply click on the "MERGE" button in the processing window.

Inspect each curve and the I Scale number which correlates to the dilutions made from the original sample. (Note: Samples at higher concentration display less noise in the tail region of the curves.)

4.7.6 P(r) Distribution

To generate the P(r) plot, load the merged data curves into PRIMUS/qt as previously described.

Load the "GNOM" module by typing "GNOM" into the command prompt area.

Click on the "GNOMPLOT" button to open a new window presenting the merged data of intensity of scattered light vs. q and pair-distance distribution function plot on the right-hand. (Note: The information on the right-hand side presents the overall quality of the pair-distance distribution function calculations.)

Adjust the data range of the merged data to avoid any significant noise at the tail end of the raw data.

Omit data points close to the beam stop in the low-q region.

To determine the D_{max}, start with a range of ~5 times the R_g obtained from the Guinier analysis. Gradually decrease this value until the P(r) plot is not abruptly dropping to zero on the Y-axis and does not have a long-tail before approaching zero.

Check that the Experimental R_g/I_0 (derived from Guinier approximation) and P(r) Rg/I_0 numbers are similar. (Note: In some cases, further manipulation on the range of data, data points, and ALPHA (a regularization parameter which tells the program the ratio of how much attention is paid to the smoothness of the distribution compared to fitting the experimental data) is also required (16) to obtain a good quality P(r) plot).

4.8 Ab initio *bead modeling and averaging*

Once the data collected at multiple concentrations are merged, or the data collected using SEC-SAXS is minimized, and the Kratky plot, P(r) plot and Guinier analysis have been verified, low-resolution structures of macromolecules and their complexes can be calculated. We have used this pipeline to study solution structures and interactions of nucleic acids, proteins, and nucleic acids-protein or protein-protein complexes (10,11,16-29). One of the most popular programs is DAMMIN, developed by Svergun (30), that is a part of the ATSAS package (31). DAMMIN employs simulated annealing protocols with preliminary input information on R_g and D_{max}. The *ab initio* modeling approaches and principles are described in detail elsewhere (32,33).

4.9 Representative results

The data analysis approach described above was utilized to calculate the R_g and D_{max} for nidogen-1, laminin γ-1 and their complex using the P(r) function. We obtained R_g values of 7.20 (±0.10) nm, 8.10 (±0.20) nm and 10.9 (±0.4) nm for nidogen-1, laminin γ-1 and their complex respectively (Figure 2 A and B). In addition, the D_{max} values of 24 nm, 26 nm, and 35 nm for nidogen-1, laminin γ-1, and their complex respectively (Figure 4.2) (10) were obtained. The DAMMIF program was used to obtain low-resolution structures of nidogen-1

and laminin γ-1 which suggested that both proteins adopt an extended shape in solution. The X and NSD values for nidogen-1 (~1 and 0.8) and laminin γ-1 (~0.9 and 0.8 respectively) were also in the acceptable range. The alignment of high-resolution structures: two domains of nidogen-1 and two of laminin γ-1 on their low-resolution structures obtained using SAXS allowed identification of their N- and C-terminal regions (10).

Nidogen-1 was identified as an interacting partner of laminin γ-139 (34), and the interaction site was mapped using X-ray crystallography to the C-terminal domains (35). However, high-resolution structures only involved interacting domains and not the full-length

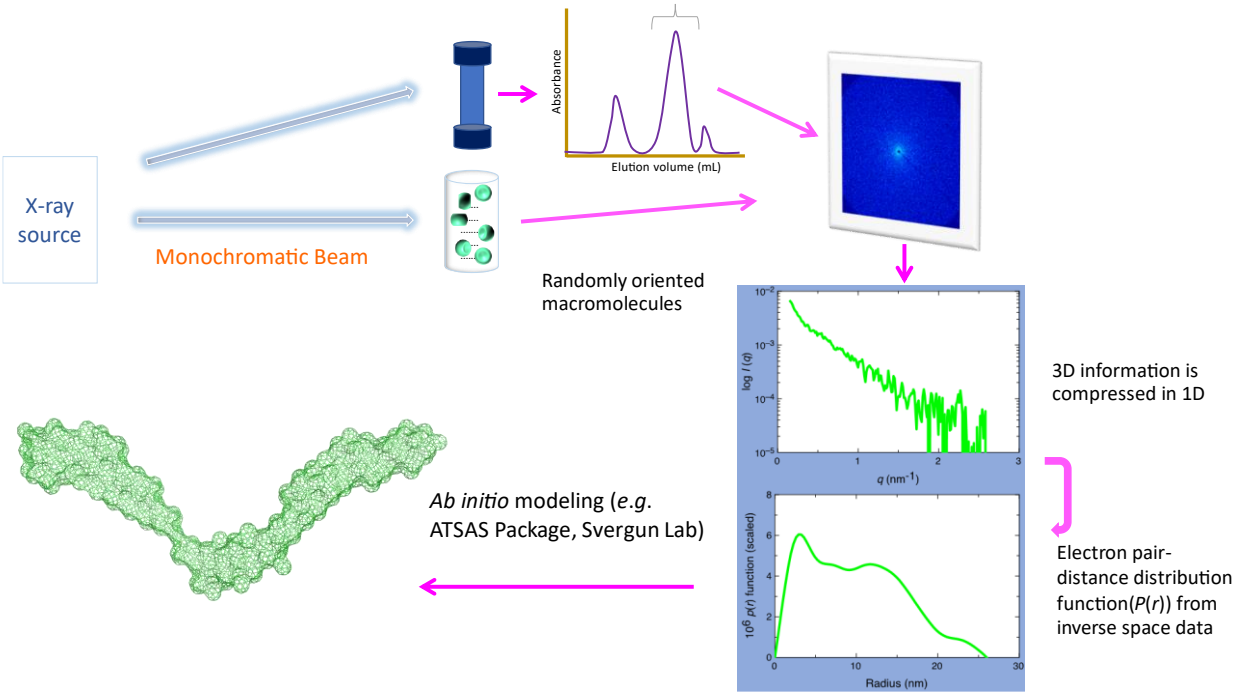


Figure 4.1. Schematics of SAXS set-up. A monodispersed preparation of biomolecules or their complexes is prepared, followed by exposure with high energy X-rays. Depending on the source (e.g., in-house vs. synchrotron), the energy of X-rays and the sample to source distance can vary. The X-rays' scattering pattern (that depends on the size and shape of biomolecules) is recorded and radially averaged to obtain a 1-dimensional plot (1D) that contains information on the intensity of scattered light with respect to the scattering angle. As buffer molecules also scatter light, the contributions from these molecules are subtracted to obtain a scattering pattern of the biomolecules of interest. At the synchrotron, prior to the SAXS data collection, an additional purification step using in-line size exclusion/high-performance chromatography is also typically performed (top view). This step is critical to remove any aggregated and/or degraded product as well as to remove any unbound biomolecules from the complex. The 1D scattering plot is converted to the electron pair-distance distribution plot (P(r) plot), which provides the radius of gyration and maximum particle dimension of biomolecules. This plot is used as the input file for the ab initio modeling packages (i.e., DAMMIN/DAMMIF) to obtain low-resolution structures of biomolecules, or other packages (i.e., SASREF/CORAL) if the high-resolution structure of parts of the biomolecules or individual biomolecules of the complex is known

nidogen-1 or the entire laminin γ-1 arm. Therefore, we purified a complex containing nidogen-1 (full length) and the laminin γ-1 arm to identify the interacting regions as well as to study the relative orientation of the N-terminal domains of both proteins. The SAXS data for the complex yielded an R_g of 10.9 (±0.4) nm and a D_{max} of 35 nm. We utilized MONSA to obtain the low-resolution structure of the entire complex which suggested that indeed, only the C-terminal region of both proteins participate in mediating interactions, whereas, the rest of the domains are far apart from each other (Figure 4.3).

4.10. Discussion

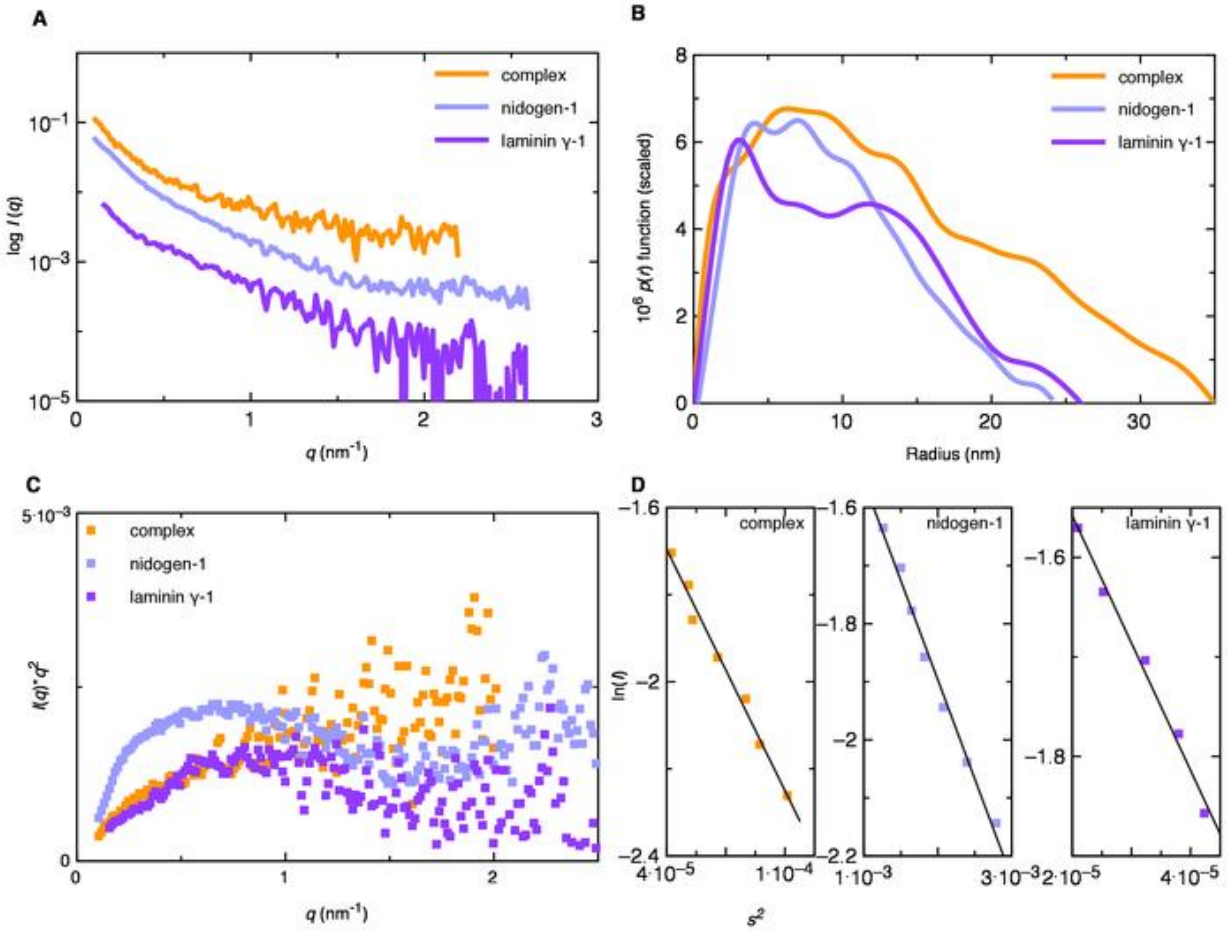


Figure 4.2. SAXS data plots (A) A plot of an intensity of scattered light *vs.* scattering angle ($q=4\pi\sin\theta/\lambda$, nm^{-1}) suggesting the quality of biomolecules (low region) and shape (high region) of biomolecules. **(B)** The electron pair-distance distribution *P*(*r*) determined from the scattering data suggest an elongated shape of biomolecules under investigation (laminin γ-1, nidogen-1, and their complex). **(C)** Kratky plot suggesting that nidogen-1 and laminin γ-1 proteins are not unfolded**. (D)** Guinier plot for nidogen-1, laminin γ-1 and their complex, indicating the linear region for determination of the radius of gyration using data at low-scattering angle.

The critical steps of SAXS data analysis outlined in the protocol section of this paper include buffer subtraction, Guinier analysis, Kratky analysis, data merging and P(r) distribution. The *ab initio* bead modeling is too extensive to be covered here in detail and is therefore only covered briefly.

At synchrotrons (e.g. DESY in Germany, Diamond in the UK and ESRF in France), it is possible to collect SAXS data for a very tiny fraction (~few µL) of each sample as the fractions are being eluted from the SEC column that is connected in-line (see Figure 4.1). The elastically scattered SAXS data is radially averaged using the packages provided by the instrument manufacturer or by the synchrotron before buffer subtraction can take place. The resulting 1D data represents the amount of scattered light (ln I(q)) on the Y-axis and scattering angle ($q=4\pi\sin\theta/\lambda$, where λ is the wavelength of incident X-rays) and is outlined in Figure 4.1. The program PRIMUS/qt (12) is used to directly subtract any background due to buffer and is described in section 4.1. Other programs such as; ScÅtter (13) (download available at www.bioisis.net) with a tutorial available at, and bioXtas RAW (14) available at https://bioxtas-raw.readthedocs.io/en/latest/index.html can be utilized as an alternative to the ATSAS package.

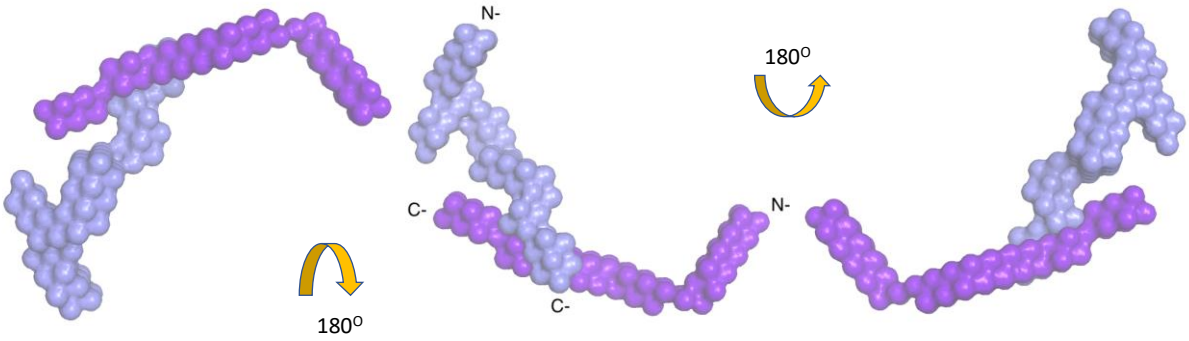


Figure 4.3. Low-resolution structure of the complex of nidogen-1, and laminin γ-1. Obtained by analysis of merged data sets using the program MONSA. The color scheme is the same as Figure 4.2.

The Guinier analysis provides information on sample aggregation and homogeneity as well as providing the radius of gyration (R_g) for the macromolecule of interest based on the SAXS data from the low q region (36). A plot is constructed with PRIMUS/qt for SAXS data obtained from each concentration, followed by curve fitting with the maximum range of up to 1.30 for $q.R_g$. A monodispersed sample preparation should provide a linear Guinier plot in this region (Figure 4.2D), whereas aggregation results in a nonlinear Guinier plot (37,38). If the Guinier analysis is linear, the degree of "unfoldedness" of a macromolecule of interest can be observed with the Kratky plot which is useful when deciding whether to perform rigid body modeling or construct ensembles of low-resolution models. A globular protein will appear in a Kratky plot to have a bell-shaped curve, whereas extended molecules or unfolded peptides will appear to plateau or even increase in the larger q range and lack the bell-shape (Figure 4.2C).

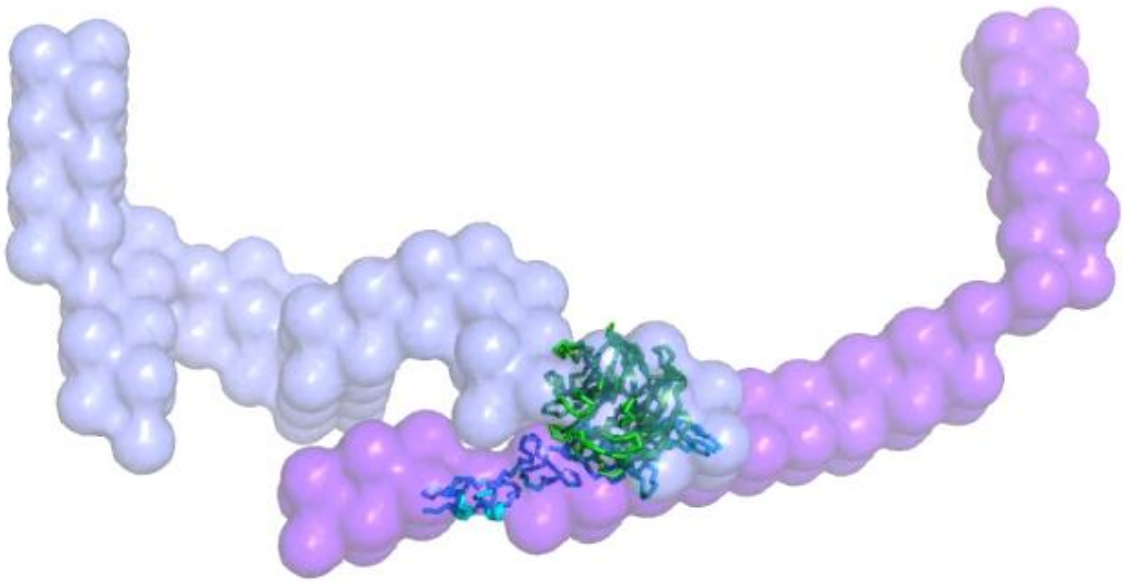


Figure 4.4 The low-resolution structure of the nidogen-1 and laminin γ-1 complex. This movie was prepared using PYMOL to visualize various structural features of the complex. The crystal structure of the laminin-nidogen complex (PDB ID: 1NPE) is shown as ribbon cartoons, highlighting the interacting sites for this complex. The color scheme is the same as Figure 2. To access this movie visit
https://www.jove.com/files/ftp_upload/58538/complex_movie.mpg

Obtaining the R_g from Guinier analysis only considers data points from the low q region of the 1D scatter plot (Figure 4.2D), however, it is possible to use almost the entire dataset to perform an indirect Fourier transformation to convert the reciprocal-space information of ln(I(q)) vs. (q) into a real space distance distribution function (P(r)) which provides information on D_{max} and R_g (Figure 4.2B) The shape of the P(r) plot represents the gross solution conformation of the macromolecule of interest (32,39). the conversion of reciprocal-space data to real-space data is a critical step but a detailed description is not within the scope of this paper. Therefore, refer to an article by Svergun (39) to understand each parameter.

Once the buffer subtracted data at individual concentrations are processed through Guinier analysis with a consistent value for R_g, followed by investigating their folding pattern using Kratky analysis, these data can be merged. The merged data for nidogen-1, laminin γ-1, and their complex were processed as described above and the resulting P(r) plots are presented in Figure 4.2B. Ideally, one should also calculate the pair-distance distribution function P(r) for each concentration to determine if SAXS data collected for each concentration provides similar R_g and D_{max} values. If the R_g and D_{max} remain similar over a wide range of concentrations, then the user should proceed. It should be noted that depending on the signal, data can be truncated prior to data merging. This is often the case if the concentrations and/or molecular weight of the macromolecules under investigation is low.

Low-resolution shape analysis using DAMMIN can be performed in various modes (e.g. Fast, Slow, Expert modes etc.). The Fast mode is an ideal first step to evaluate if the P(r) plot provides good quality models. Typically, at least 10 models should be obtained for each P(r) plot to check if reproducible results, in terms of the low-resolution structure, are obtained, with a low goodness of fit parameter called χ (a value of 0.5-1.0 is considered good based on our extensive work), a value that describes an agreement between experimentally collected SAXS data and model-derived data. For publication purpose, we typically use Slow or Expert

mode and calculate at least 15 models. In addition to DAMMIN, a faster version of it, DAMMIF (40) as well as GASBOR (41) are also alternatives. Furthermore, to study protein-protein or protein-nucleic acid complexes, it is possible to use the MONSA program (29) which facilitates simultaneous fitting of the individual SAXS data for both macromolecules as well as their complex. For more details on high-resolution model calculations as well for RNA-protein interaction studies, refer to a recent article by Patel et al (2).

SAXS is theoretically simple but undoubtedly a highly complementary method to other structural biology tools and results in low-resolution structural data that can be used on its own or in conjunction with high-resolution techniques to elucidate information about macromolecular structure and dynamics. As long as a monodispersed preparation of macromolecules and their complexes can be obtained, SAXS can be utilized to study in-solution structure and interactions of any type of biological macromolecule. In the case of the complex discussed here, it is remarkable that less than 10% of the overall accessible surface area of nitrogen-1 and laminin γ-1 is buried in this complex, whereas the rest of domains of both proteins are freely accessible to interact with other proteins at the extracellular matrix to maintain its structural rigidity (Figure 4.3). Obtaining such information for a complex with ~240 kDa would be very challenging using other structural biology techniques such as XRC, NMR, and CEM.

4.11 Conclusions

Uncovering protein structure via X-ray crystallography or NMR is an inherently time-consuming process. This bottleneck in structure determination is one area where SAXS shows its strength as a structural technique; data acquisition for a single SAXS experiment can take less than an hour and with the help of streamlined analysis software, analysis can be done quickly, and efficiently. SAXS has the potential to greatly increase throughput of structural studies as a stand-alone technique because it offers a low-resolution model of the macromolecular structure before high-resolution data is available. A barrier to other structural

techniques is the requirement for a highly pure, concentrated sample for data acquisition, which necessitates a high level of protein expression and stability over a long period of time. While SAXS samples also need to be pure and concentrated, the sample volumes are roughly 100 µL making SAXS a relatively inexpensive method of analysis compared to other structural techniques. Moreover, SAXS coupled with size exclusion chromatography is becoming increasingly common which provides an additional quality control step. Recently there has been strong advances in the combination of NMR and SAXS data using the Ensemble Optimization Method (EOM) (42,43) to elucidate flexible systems. A recent paper by Mertens and Svergun (44) the authors describe multiple recent examples of EOM SAXS in combination with NMR, along with many other examples of SAXS data being used in conjunction with NMR. Advances are continually being made in the field of SAXS, and new techniques are being developed for SAXS to be used in conjunction, not just complimentary, to other structural techniques. Consequently, we believe that the demand for SAXS will only increase over time, especially in conjunction with NMR to characterize dynamic systems where functions are defined by flexibility.

4.12 Data Availability

The Movie associated with this publication is open access and can be found at https://www.jove.com/v/58538/structural-studies-of-macromolecules-in-solution-using-small-angle-x--ray-scattering, in addition to the movie transcript.

4.13 References

1. Svergun, D.I. and Koch, M.H. (2002) Advances in structure analysis using small-angle scattering in solution. *Current opinion in structural biology*, **12**, 654-660.
2. Patel, T.R., Chojnowski, G., Koul, A., McKenna, S.A. and Bujnicki, J.M. (2017) Structural studies of RNA-protein complexes: A hybrid approach involving hydrodynamics, scattering, and computational methods. *Methods*, **118**, 146-162.
3. Stetefeld, J., McKenna, S.A. and Patel, T.R. (2016) Dynamic light scattering: a practical guide and applications in biomedical sciences. *Biophysical reviews*, **8**, 409-427.
4. Uversky, V.N., Gillespie, J.R. and Fink, A.L. (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins: structure, function, and bioinformatics*, **41**, 415-427.
5. Uversky, V.N., Gillespie, J.R., Millett, I.S., Khodyakova, A.V., Vasiliev, A.M., Chernovskaya, T.V., Vasilenko, R.N., Kozlovskaya, G.D., Dolgikh, D.A. and Fink, A.L. (1999) Natively unfolded human prothymosin α adopts partially folded collapsed conformation at acidic pH. *Biochemistry*, **38**, 15009-15016.
6. Kikhney, A.G. and Svergun, D.I. (2015) A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS letters*, **589**, 2570-2577.
7. FoXS, F. (2016) MultiFoXS: single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles; D Schneidman-Duhovny, M Hammel, JA Tainer, A Sali. *Nucleic acids research*, W424-W429.
8. Svergun, D., Barberato, C. and Koch, M.H. (1995) CRYSOL–a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of applied crystallography*, **28**, 768-773.
9. Pérez, J. and Vachette, P. (2017) A successful combination: coupling SE-HPLC with SAXS. *Biological small angle scattering: techniques, strategies and tips*, 183-199.
10. Patel, T.R., Bernards, C., Meier, M., McEleney, K., Winzor, D.J., Koch, M. and Stetefeld, J. (2014) Structural elucidation of full-length nidogen and the laminin–nidogen complex in solution. *Matrix Biology*, **33**, 60-67.
11. Meier, M., Moya-Torres, A., Krahn, N.J., McDougall, M.D., Orriss, G.L., McRae, E.K.S., Booy, E.P., McEleney, K., Patel, T.R. and McKenna, S.A. (2018) Structure and hydrodynamics of a DNA G-quadruplex with a cytosine bulge. *Nucleic acids research*, **46**, 5319-5331.
12. Franke, D., Petoukhov, M., Konarev, P., Panjkovich, A., Tuukkanen, A., Mertens, H., Kikhney, A., Hajizadeh, N., Franklin, J. and Jeffries, C. (2017) ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *Journal of applied crystallography*, **50**, 1212-1225.
13. Forster, S., Apostol, L. and Bras, W. (2010) Scatter: software for the analysis of nano- and mesoscale small-angle scattering. *Journal of Applied Crystallography*, **43**, 639-646.
14. Hopkins, J.B., Gillilan, R.E. and Skou, S. (2017) BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *J Appl Crystallogr*, **50**, 1545-1553.
15. Putnam, C.D., Hammel, M., Hura, G.L. and Tainer, J.A. (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly reviews of biophysics*, **40**, 191-285.

16. Patel, T.R., Reuten, R., Xiong, S., Meier, M., Winzor, D.J., Koch, M. and Stetefeld, J. (2012) Determination of a molecular shape for netrin-4 from hydrodynamic and small angle X-ray scattering measurements. *Matrix Biology*, **31**, 135-140.
17. Dzananovic, E., Patel, T.R., Chojnowski, G., Boniecki, M.J., Deo, S., McEleney, K., Harding, S.E., Bujnicki, J.M. and McKenna, S.A. (2014) Solution conformation of adenovirus virus associated RNA-I and its interaction with PKR. *Journal of structural biology*, **185**, 48-57.
18. Meier, M., Patel, T.R., Booy, E.P., Marushchak, O., Okun, N., Deo, S., Howard, R., McEleney, K., Harding, S.E. and Stetefeld, J. (2013) Binding of G-quadruplexes to the N-terminal recognition domain of the RNA helicase associated with AU-rich element (RHAU). *Journal of Biological Chemistry*, **288**, 35014-35027.
19. Dzananovic, E., Patel, T.R., Deo, S., McEleney, K., Stetefeld, J. and McKenna, S.A. (2013) Recognition of viral RNA stem–loops by the tandem double-stranded RNA binding domains of PKR. *Rna*, **19**, 333-344.
20. Bacik, J.-P., Tavassoli, M., Patel, T.R., McKenna, S.A., Vocadlo, D.J., Khajehpour, M. and Mark, B.L. (2014) Conformational itinerary of Pseudomonas aeruginosa 1, 6-anhydro-N-acetylmuramic acid kinase during its catalytic cycle. *Journal of Biological Chemistry*, **289**, 4504-4514.
21. Deo, S., Patel, T.R., Chojnowski, G., Koul, A., Dzananovic, E., McEleney, K., Bujnicki, J.M. and McKenna, S.A. (2015) Characterization of the termini of the West Nile virus genome and their interactions with the small isoform of the 2′ 5′-oligoadenylate synthetase family. *Journal of structural biology*, **190**, 236-249.
22. Deo, S., Patel, T.R., Dzananovic, E., Booy, E.P., Zeid, K., McEleney, K., Harding, S.E. and McKenna, S.A. (2014) Activation of 2′ 5′-oligoadenylate synthetase by stem loops at the 5′-end of the West Nile virus genome. *PloS one*, **9**, e92545.
23. Vadlamani, G., Thomas, M.D., Patel, T.R., Donald, L.J., Reeve, T.M., Stetefeld, J., Standing, K.G., Vocadlo, D.J. and Mark, B.L. (2015) The β-lactamase gene regulator AmpR is a tetramer that recognizes and binds the D-Ala-D-Ala motif of its repressor UDP-N-acetylmuramic acid (MurNAc)-pentapeptide. *Journal of biological chemistry*, **290**, 2630-2643.
24. Ariyo, E.O., Booy, E.P., Patel, T.R., Dzananovic, E., McRae, E.K., Meier, M., McEleney, K., Stetefeld, J. and McKenna, S.A. (2015) Biophysical characterization of G-quadruplex recognition in the PITX1 mRNA by the specificity domain of the helicase RHAU. *PloS one*, **10**, e0144510.
25. Hyde, E.I., Callow, P., Rajasekar, K.V., Timmins, P., Patel, T.R., Siligardi, G., Hussain, R., White, S.A., Thomas, C.M. and Scott, D.J. (2017) Intrinsic disorder in the partitioning protein KorB persists after co-operative complex formation with operator DNA and KorA. *Biochemical Journal*, **474**, 3121-3135.
26. Dzananovic, E., Chojnowski, G., Deo, S., Booy, E.P., Padilla-Meier, P., McEleney, K., Bujnicki, J.M., Patel, T.R. and McKenna, S.A. (2017) Impact of the structural integrity of the three-way junction of adenovirus VAI RNA on PKR inhibition. *PloS one*, **12**, e0186849.
27. Krahn, N., Meier, M., To, V., Booy, E.P., McEleney, K., O'Neil, J.D., McKenna, S.A., Patel, T.R. and Stetefeld, J. (2017) Nanoscale assembly of high-mobility group AT-Hook 2 protein with DNA replication fork. *Biophysical journal*, **113**, 2609-2620.
28. Patel, T.R., Besong, T.M., Meier, M., McEleney, K., Harding, S.E., Winzor, D.J. and Stetefeld, J. (2018) Interaction studies of a protein and carbohydrate system using an integrated approach: a case study of the miniagrin–heparin system. *European Biophysics Journal*, **47**, 751-759.
29. Svergun, D.I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophysical journal*, **76**, 2879-2886.
30. Blanchet, C.E. and Svergun, D.I. (2013) Small-angle X-ray scattering on biological macromolecules and nanocomposites in solution. *Annual review of physical chemistry*, **64**, 37-54.

31. Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H. and Svergun, D.I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of applied crystallography*, **36**, 1277-1282.
32. Petoukhov, M.V. and Svergun, D.I. (2013) Applications of small-angle X-ray scattering to biomacromolecular solutions. *The international journal of biochemistry & cell biology*, **45**, 429-437.
33. Volkov, V.V. and Svergun, D.I. (2003) Uniqueness of ab initio shape determination in small-angle scattering. *Journal of applied crystallography*, **36**, 860-864.
34. Baumgartner, R., Czisch, M., Mayer, U., Pöschl, E., Huber, R., Timpl, R. and Holak, T.A. (1996) Structure of the Nidogen Binding LE Module of the Laminin γ1 Chain in Solution. *Journal of Molecular Biology*, **257**, 658-668.
35. Stetefeld, J., Mayer, U., Timpl, R. and Huber, R. (1996) Crystal structure of three consecutive laminin-type epidermal growth factor-like (LE) modules of laminin gamma1 chain harboring the nidogen binding site. *J Mol Biol*, **257**, 644-657.
36. Tuukkanen, A.T. and Svergun, D.I. (2014) Weak protein–ligand interactions studied by small-angle X-ray scattering. *The FEBS journal*, **281**, 1974-1987.
37. Koch, M.H., Vachette, P. and Svergun, D.I. (2003) Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Quarterly reviews of biophysics*, **36**, 147-227.
38. Rambo, R.P. and Tainer, J.A. (2013) Accurate assessment of mass, models and resolution by small-angle scattering. *Nature*, **496**, 477-481.
39. Svergun, D. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of applied crystallography*, **25**, 495-503.
40. Kozin, M.B. and Svergun, D.I. (2001) Automated matching of high-and low-resolution structural models. *Journal of applied crystallography*, **34**, 33-41.
41. Franke, D. and Svergun, D.I. (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *Journal of applied crystallography*, **42**, 342-346.
42. Bernadó, P., Mylonas, E., Petoukhov, M.V., Blackledge, M. and Svergun, D.I. (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc*, **129**, 5656-5664.
43. Tria, G., Mertens, H.D., Kachala, M. and Svergun, D.I. (2015) Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ*, **2**, 207-217.
44. Mertens, H.D.T. and Svergun, D.I. (2017) Combining NMR and small angle X-ray scattering for the study of biomolecular structure and dynamics. *Arch Biochem Biophys*, **628**, 33-41.

Chapter 5. Analytical Ultracentrifugation

5.1 Foreword

Chapters 2-5 are written because of my extensive focus on learning and applying numerous biophysical techniques throughout my Ph.D. While this is not an extensive list, I primarily focused on four techniques, MST, SEC-MALS, SAXS, and AUC. This chapter focuses on AUC, the theory and physics behind the technique, applications of the technique, experimental considerations, and personal applications for RNA samples in AUC. AUC experiments were performed using AUC instruments at the Canadian Center for Hydrodynamics at the University of Lethbridge and with help from Amy Henrickson and Dr. Borries Demeler.

5.2 Abstract

Analytical ultracentrifugation is a powerful biophysical technique invent over 100 years ago. By combining modern optical systems with extremely powerful centrifugal force, AUC can provide evidence of the size, shape, and density of molecules in solution and easily distinguish between different species in a heterogeneous mixture.

Analysis of solutions of molecules is of utmost importance when purifying biological molecules from native (or synthetic) sources. While other techniques exist within the same space (SEC-MALS, dynamic light scattering), AUC can modulate its resolution by changing the centrifugal force (speed of the rotor). This modulation allows AUC to perform an incredibly diverse number of experiments, especially considering that AUC can be performed as either a sedimentation velocity (SV-AUC), or sedimentation equilibrium (SE-AUC). Here, AUC is described as it relates to the characterization of biomolecules in solution and specifically comment on my experiences with AUC on RNA samples.

5.3 Theory

In an AUC experiment, a sample in solution is exposed to a large centrifugal force, a term coined by Christiaan Huygens in 1659 and roughly translated from Latin meaning "fleeing the center." It was not until 200 years later that Antonin Prandtl used the idea of centrifugation to speed up the process by which cream separates from milk. Five years later, a Swiss biologist, Friedrich Miescher, used a centrifuge to isolate "nuclein" from cells, the biological material we know today as DNA (1). However, it was not until Theodor Svedberg, in 1923, invented the first ultracentrifuge, which was used as an analytical instrument (2). Subsequently, Svedberg was awarded the Nobel Prize in Chemistry in 1926. However, often, it is miscredited that the Nobel Prize was for his invention of the ultracentrifuge when in fact, it was for his work in colloid chemistry. Svedberg's work which earned him the Nobel Prize effectively proved that macromolecules comprise covalently bonded atoms. Svedberg and colleagues' pioneering ultracentrifuge could achieve ~42,000 rpm or 900,000 xg. Svedberg's contributions earned him the Nobel Prize and the term Svedberg unit, often represented as S (sometimes Sv). A Svedberg unit is a non-SI measurement for the speed with which a macromolecule will sediment in solution, with one Svedberg unit being equal to (10^{-13} s). The Svedberg is used to define the sedimentation coefficient, which is the ratio between the speed of a macromolecule in solution in a centrifuge to its acceleration. Therefore, in its base terms, a sedimentation coefficient is the time a particle takes to achieve its terminal velocity under a given acceleration (assuming no counterforces).

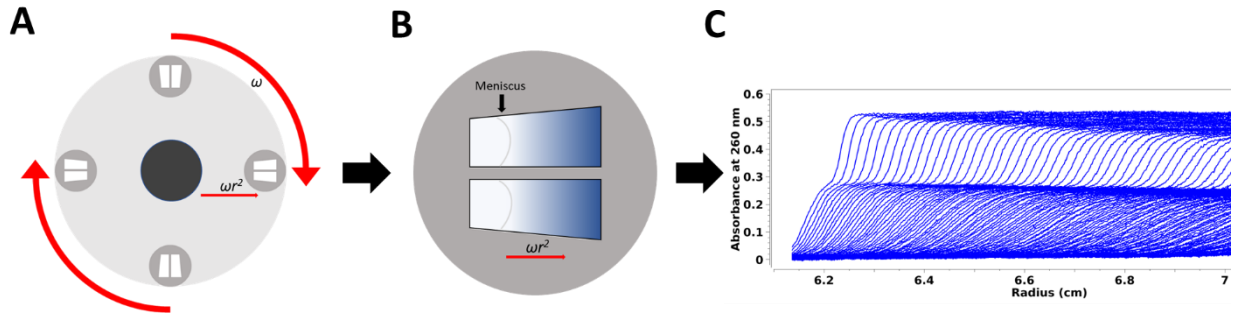


Figure 5.1. Schematic of analytical ultracentrifugation going from centrifugal force to usable data. A) Overhead view of a rotor schematic showing rotation and resulting centrifugal force. **B)** Representation of a typical AUC cell showing a change in concentration as a function of cell position (gradient) because of centrifugal force. **C)** Example AUC scan data showing concentration (absorbance) as a function of cell position.

AUC, in essence, applies a centrifugal force to a solution of macromolecules (Figure 5.1A), which, if sufficient, creates a concentration boundary that moves away from the centrifugal force (Figure 5.1B). The resulting concentration of solute in the cell is measured by UV optics (absorbance) and plotted as a function of concentration vs. radius (position relative to meniscus), and each scan (blue line) is a new timepoint (Figure 5.1C). This measurement allows the application of the Svedberg equation (3) (equation 5.1) to determine the sedimentation coefficient.

$$s = \frac{V_r}{\omega^2 r} = \frac{M_w(1 - \bar{v}\rho)}{N_A f_o}$$

(5.1)

In equation 5.1, the sedimentation coefficient is represented by *s*, while V_r represents the radial velocity. Angular velocity is represented by ω, with *r* being the radial position (of the solvent boundary). M_w represents molecular weight, and partial specific volume is represented by vbar ($\bar{v}$), with the density of the solvent represented by rho (ρ). N_A is Avagadro's number, and f_o is the frictional coefficient. The Svedberg equation requires some additional input from other equations, such as the Stokes equation (4), to calculate the frictional coefficient f_o (equation 5.2)

$$f_o = 6\pi\eta R_o$$

(5.2)

In the above equation, eta (η) is representative of the viscosity of the solution, while R_o is the radius of a sphere. One can then combine the Svedberg and Stokes equation (5,6) with the radius of a molecular sphere and calculate the sedimentation coefficient for spherical molecules in water (at 20°C). This correction to water and 20°C is necessary to standardize measurements done under different buffer conditions and performed on different machines in different laboratories. *W* represents the notation for water, while *B* represents the notation for a buffer for the below equation (5.3) (7).

$$s_{20,w} = s_{T,B}\left(\frac{\eta_{T,B}}{\eta_{20,W}}\right)\frac{(1-\ \bar{v}\rho\ )_{20,W}}{(1-\ \bar{v}\rho)_{T,B}}$$

(5.3)

Additionally, sedimentation equilibrium experiments (8-12) are also possible thanks to the Lamm equation (13), which can be applied to a single sedimenting species in a sector-shaped cell (equation 5.4).

$$\frac{dc}{dt} = \left(\frac{1}{r}\right)\left(\frac{d}{dr}\right)\left[rD(M)\frac{dc}{dr} - s(M)\omega^2 r^2 c\right]$$

(5.4)

Solute concentration, time, and radius (of the concentration boundary) are represented by *c*, *t,* and *r*. *D(M)* represents the diffusion of the molecule, while *s(M)* represents the sedimentation coefficient of the molecule. Sedimentation and diffusion are opposing forces, with sedimentation trying to concentrate the solute towards the outer radius of the AUC cell and diffusion attempting to equalize the solute throughout the cell. Therefore, by varying the centrifugal force in a sedimentation equilibrium experiment, sedimentation and diffusion are equalized. This equalization permits a considerable simplification and allows for the rearrangement of equation 5.4, resulting in equation 5.5:

$$M_w(r) = \frac{\partial \ln J(r)}{\partial(r^2)}\frac{2RT}{(1-\bar{v}\rho)\omega^2}$$

(5.5)

In this simplification, $J(r) = J(a) + j(r)$, where $J(r)$ represents the absolute concentration at a radius of *r*, which is proportional to the concentration at the meniscus *j(r)*, and the absolute meniscus concentration *J(a)*. This equation allows us to plot $\ln c\ vs.\ r^2$ for the case of a UV optics system, where the slope of the equation will be $M_w(1-\ \bar{v}\rho)\omega^2/RT$ allowing for the calculation of molecular weight (14,15). Notably, the molecular weight obtained from

SE experiments is an average of all molecules in solution, so it should ideally only be performed on monodispersed samples. Additionally, the molecular weight from SE-AUC is the *apparent* molecular weight ($M_{w,\ app}$), this distinction is made because of the measurement occurring at a specific concentration. If the same experiment is performed at multiple concentrations and extrapolated to zero concentration, then the derived molecular weight is the *ideal* molecular weight.

If the molecular mass of the molecule of interest is known, in addition to the partial specific volume ($\bar{v}$), then calculating the diffusion coefficient is also possible using SE. If one applies the Stokes-Einstein equation (5.2) to the Svedberg equation (5.1), the simplified result is equation 5.6:

$$D = \frac{k_B T}{6\pi\eta_0\eta_r \left(\frac{f}{f_o}\right) R(M_w, \bar{v})} \tag{5.6}$$

The diffusion coefficient of the molecule in question is denoted by *D*, while k_B represents Boltzmann's constant (1.380×10^{-23} J·K^{-1}). The viscosity of the standard solution and relative viscosity of the experimental solution is denoted by η_0 and η_r, respectively. Finally, *R* represents the volume of a theoretical sphere of the same volume as the molecule of interest (16). Taking all these theoretical considerations and equations together means that AUC is a robust and powerful method for the characterization of biomolecular compounds in solution.

5.4 Applications

AUC is widely applied to biomolecular science to readily study homogeneity, molecular weight, oligomerization (aggregation), stability, and interactions (17-19). Modern AUC experiments are be extremely versatile and can be executed in various conditions, including wide ranges of pH, ionic concentration, temperature (Beckman Coulter Optima AUC

temperature range is from 0 to 40° C) (20), and solute concentrations. Another phenomenal property of AUC is that all the hydrodynamic property determinations (s, M_w, D, $\bar{v}$) do not require calibration with a known standard like other hydrodynamic techniques (21,22). As previously mentioned, AUC is primarily performed in two distinct ways: sedimentation velocity and sedimentation equilibrium. Each method has benefits and limitations, and it is vital to choose the correct technique for the system of study.

Sedimentation velocity experiments can provide information on the size, shape, interactions, and in some cases, conformational changes of macromolecules in solution. One of the primary uses for SV-AUC experiments is to evaluate the purity and state of purified biological molecules (23-26). Typical SV-AUC experiments are represented by sedimentation coefficient vs. change in concentration as a function of time; for example, figure 5.2 shows the sedimentation profile of BSA, where each oligomeric species is distinctly present (monomer, dimer, and tetramer). Often, SEC is used as a technique to validate the purity of macromolecules in solution, but resolution can be an issue. AUC has the benefit of much higher resolution, even being capable of discerning between different RNAs of the same molecular weight based on tertiary folding (25,27). SV-AUC has been used for decades to

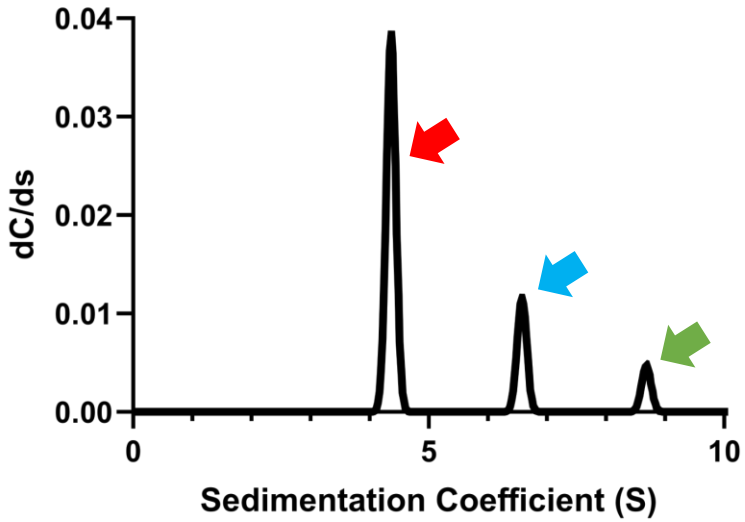


Figure 5.2. Sedimentation velocity analysis of BSA. Red, blue, and green arrows represent BSA monomer, dimer, and tetramer respectively. 0.45 OD (280 nm) of sample was run at 45k rpm for 11 hours 40 min. Data contributed by Amy Henrickson.

characterize systems, most famously the ribosomal subunits (28). Some recent studies using AUC include: investigating the binding cooperativity and affinity of NanR to its DNA substrate and its subsequent DNA-protein oligomeric complex (29), characterizing the molecular weight distribution of the BRPF1 bromodomain in complex with mono- and di-acetylated histone ligands (30), measuring the nucleic acid loading profile of lipid nanoparticles for use in gene therapeutics (31), and analyzing the stoichiometry between DM64 and myotoxin II through sedimentation and molecular weight comparisons (32). A downside to SV-AUC is the time commitment; a single AUC experiment can be very time-consuming, often running for 12 hours, like the example provided in Figure 5.2, to determine sample quality. Whereas, molecular weight determination compared to other techniques such as SEC-MALS, only requires 30 min, and the new technique mass-photometry (33), can be performed in a few minutes according to Refeyn Ltd., the creators of the instrument, makes other techniques for M_w determination, more enticing than AUC.

Sedimentation equilibrium experiments, in comparison to SV-AUC, are performed under considerably less centrifugal force. This reduction in centrifugal force causes an equilibration between the sedimentation and the diffusion, resulting in a specific-shaped concentration gradient within the cell (22,34,35). This equilibration changes the method from a hydrodynamic to a thermodynamic molecular weight measurement in solution (8,10,11). SE-AUC has been used successfully to gain information on stoichiometry, assemblies of macromolecules, as well as complexes between proteins, and proteins and nucleic acids (9,36), and to gain thermodynamic measurements of weak protein self-interactions (37). Some recent uses of SE-AUC include: studying a self-associating glycopeptide antibiotic in aqueous conditions (38), comparing hydrodynamic properties of butvar B-98 and PDMS-OH, two non-aqueous soluble archaeological wood consolidants (39), investigating how human mitochondrial AAA+ ATPase SKD3 assembles into oligomers at high protein concentration in the presence of adenine (40), to investigate phase separation of binary mixtures induced by

soft centrifugal fields (41), and the analysis of the weak oligomerization of CBM_{CipA} from *Clostridium thermocellum* (42). SE-AUC, while a useful technique, has been outshined in recent years with the invention and application of other biophysical techniques. Often, SE-AUC experiments require over 24 hours to perform, with other techniques offering hydrodynamic or thermodynamic calculations in far less time. SEC-MALS offers absolute molecular weight determination in ~30 min, mass photometry (mentioned above) provides molecular weight in less than 5 min, while dynamic light scattering also provides diffusion coefficient measurements in ~30 min. Isothermal titration calorimetry (ITC) can alternatively provide information about binding affinities of biomolecular interactions (43). Nonetheless, SE-AUC remains a strong method for the characterization of biomolecules.

5.5 Considerations and Personal Applications

All considerations are regarding my experience performing SV-AUC on RNA samples throughout my graduate studies. Many biophysical techniques are aimed toward the characterization of proteins, but there have been some studies of RNA, specifically RNA tertiary structure (44-47), using SV-AUC. I believe that AUC can be a powerful technique for studying RNA tertiary structure and could be employed more.

As with virtually any biophysical technique, the quality of any input product directly reflects the quality of the output data. RNA is particularly fickle, influenced by virtually any solution condition, including but not limited to; salt concentration, freeze-thaw cycles, pH, and temperature. Virtually every molecular biology company sells an *in vitro* transcription and purification kit, which claims to give you "pure" RNA. From my experience, however, there is a significant difference between "pure" RNA and high-quality RNA suitable for biophysical experiments. A problem is the standard quality control step for RNA, which is denaturing gel electrophoresis. Denaturing electrophoresis will allow you to assess the length of transcribed RNA qualitatively. Therefore, simply using denaturing electrophoresis will not provide any information about the "foldedness" of your RNA of interest. AUC, however, is not run under

denaturing conditions (though it can be) (24), but it importantly gives some information on the hydrodynamics of the RNA molecule, which is dependent on tertiary structure. Performing hydrodynamic experiments means that considerations must be addressed to ensure any experiment contains "correctly" folded RNA. The most widely used technique to "fold" RNA is the "snap-cool" or "heat-cool" method, which involves heating the RNA to 95°C, then either cooling it on ice or cooling gradually at room temperature. Alternatively, there is the belief that RNA transcribed from a polymerase (like T7 RNA polymerase) produces RNA that co-transcriptionally folds into a tertiary structure. This difference was investigated via AUC in Figure 5.3, which shows the difference between T7-produced 570 nucleotide RNA, after purification via SEC (figure 5.3A), that was either co-transcriptionally folded or heat-cooled. Co-transcriptionally folded RNA shows a primary peak at ~10.5 S with a minor right shoulder and an additional peak at ~15.5 S (figure 5.3B red), while heat-cooled RNA displays a drastic change, showing a bi-modal distributed peak at ~10.5 S and 12 S, and an additional bi-modal peak at ~14 – 15.5 S (figure 5.3B black). The significant change in the RNA after heat cooling is hard to ignore, yet this " refolding " is used consistently throughout RNA science. Likely, "refolding" could be suitable for shorter RNA (<100 nt), but I would argue it should be

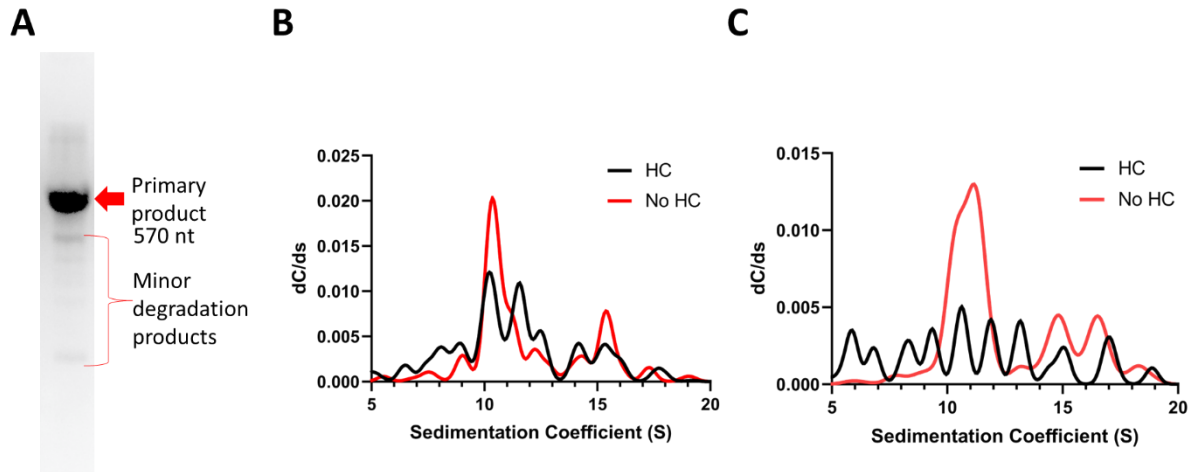


Figure 5.3. Sedimentation velocity experiments highlighting differences between purification conditions of a 570 nucleotide RNA. **A)** Urea PAGE analysis of a 570 nt long RNA product produced via *in vitro* transcription and purified using SEC (Superdex 400). **B)** Sedimentation velocity AUC of the 570 nt long RNA from panel A in the presence of 10 mM $MgCl_2$, either heat-cooled (HC) or not heat-cooled, meaning co-transcriptionally folded (no HC). **C)** Sedimentation velocity AUC of the 570 nt long RNA from panel A in the presence of 20 mM $MgCl_2$, either heat-cooled, or not. AUC experiments were performed at 28,000 rpm for 16 hours.

investigated on a case-by-case basis for lncRNA very carefully, and only used once proper quality control, like AUC has been performed.

Magnesium concentration can also be a potential modifier of RNA tertiary structure. Figure 5.3C shows an experiment performed on the same 570 nt RNA as in Figure 5.3B; the only difference is that the magnesium concentration is changed from 10 mM to 20 mM. The co-transcriptionally folded RNA shows a prominent, broad peak at ~11 S and another bi-modal peak at ~15 and 16.5 S. However, upon heat cooling the same RNA, in 20 mM magnesium chloride, the prominent peak at ~11 S is removed, and multiple peaks from 5.5 – 19 S are seen (Figure 5.3C). Magnesium ions are a known co-factor for RNA folding (48), and heat-cooling RNA in the presence of different concentrations of Mg^{2+} can drastically alter the structure of an RNA, especially long RNA. Each lncRNA will typically behave differently under different concentrations of Mg^{2+} (46), and this effect should also be investigated on a case-by-case basis. Therefore, each lncRNA should be treated as a unique entity, like a recombinantly purified protein, with a carefully optimized protocol, and investigated as such.

Overall, AUC is an effective technique with unparalleled resolution when investigating the biophysical properties of molecules in solution. Sedimentation coefficient analysis can provide invaluable information about the folding properties of RNA in solution; however, SV-AUC is often overlooked as a tool to study RNA, likely because of the barriers of entry to the technique (expertise and cost). Ultimately, if access to an AUC is possible, I believe it should be incorporated into any RNA researchers' repertoire.

5.6 References

1. Dahm, R. (2005) Friedrich Miescher and the discovery of DNA. *Developmental biology*, **278**, 274-288.
2. Koehler, C.S.W. (2003) Devloping the Ultracentrifuge. *Today's Chemist at Work*, 63-66.
3. Philpot, J.S.L. (1940). Nature Publishing Group UK London.
4. Miller, C.C. (1924) 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*, **106**, 724-749.
5. Ho, P.S., Van Holde, K., Johnson, W. and Shing, P. (1998) *Principles of physical biochemistry*. Upper Saddle River.
6. Swanson, E., Teller, D.C. and de Haën, C. (1980) Creeping flow translational resistance of rigid assemblies of spheres. *The Journal of Chemical Physics*, **72**, 1623-1628.
7. Ralston, G.B. (1993) *Introduction to analytical ultracentrifugation*. Beckman California:.
8. Harding, S.E. (1994) Determination of absolute molecular weights using sedimentation equilibrium analytical ultracentrifugation. *Microscopy, Optical Spectroscopy, and Macroscopic Techniques*, 75-84.
9. Pearce, T.C., Rowe, A.J. and Turnock, G. (1975) Determination of the molecular weights of RNAs by low-speed sedimentation equilibrium: 16 S ribosomal RNA as a model compound. *Journal of Molecular Biology*, **97**, 193-205.
10. Harding, S.E., Berth, G., Ball, A., Mitchell, J.R. and de la Torre, J.G. (1991) The molecular weight distribution and conformation of citrus pectins in solution studied by hydrodynamics. *Carbohydrate Polymers*, **16**, 1-15.
11. Gillis, R.B., Adams, G.G., Besong, D.T., Machová, E., Ebringerová, A., Rowe, A.J., Harding, S.E. and Patel, T.R. (2017) Application of novel analytical ultracentrifuge analysis to solutions of fungal mannans. *European Biophysics Journal*, **46**, 235-245.
12. Chittuluru, J.R., Chaban, Y., Monnet-Saksouk, J., Carrozza, M.J., Sapountzi, V., Selleck, W., Huang, J., Utley, R.T., Cramet, M. and Allard, S. (2011) Structure and nucleosome interaction of the yeast NuA4 and Piccolo–NuA4 histone acetyltransferase complexes. *Nature structural & molecular biology*, **18**, 1196-1203.
13. Lamm, O. (1929) *Die differentialgleichung der ultrazentrifugierung*. Almqvist & Wiksell.
14. Harding, S.E., Vårum, K., Stokke, B.r.T. and Smidsrød, O. (1991) Molecular weight determination of polysaccharides. *Advances in carbohydrate analysis*, **1**, 63-144.
15. Patel, T.R., Chojnowski, G., Koul, A., McKenna, S.A. and Bujnicki, J.M. (2017) Structural studies of RNA-protein complexes: A hybrid approach involving hydrodynamics, scattering, and computational methods. *Methods*, **118**, 146-162.
16. Laue, T.M. (1992) Computer-aided interpretation of analytical sedimentation data for proteins. *Analytical ultracentrifugation in biochemistry and polymer science.*, 90-125.
17. Meier, M., Patel, T.R., Booy, E.P., Marushchak, O., Okun, N., Deo, S., Howard, R., McEleney, K., Harding, S.E. and Stetefeld, J. (2013) Binding of G-quadruplexes to the N-terminal recognition domain of the RNA helicase associated with AU-rich element (RHAU). *Journal of Biological Chemistry*, **288**, 35014-35027.
18. Dzananovic, E., Patel, T.R., Chojnowski, G., Boniecki, M.J., Deo, S., McEleney, K., Harding, S.E., Bujnicki, J.M. and McKenna, S.A. (2014) Solution conformation of adenovirus virus associated RNA-I and its interaction with PKR. *Journal of structural biology*, **185**, 48-57.
19. Greive, S.J., Lins, A.F. and von Hippel, P.H. (2005) Assembly of an RNA-protein complex: binding of NusB and NUsE (S10) proteins to boxA RNA nucleates the formation of the antitermination

complex involved in controlling rRNA transcription in Escherichia coli. *Journal of Biological Chemistry*, **280**, 36397-36408.
20. Edwards, G.B., Muthurajan, U.M., Bowerman, S. and Luger, K. (2020) Analytical Ultracentrifugation (AUC): An Overview of the Application of Fluorescence and Absorbance AUC to the Study of Biological Macromolecules. *Current Protocols in Molecular Biology*, **133**, e131.
21. Laue, T.M. and Stafford III, W.F. (1999) Modern applications of analytical ultracentrifugation. *Annual review of biophysics and biomolecular structure*, **28**, 75-100.
22. Fujita, H. (1975) *Foundations of ultracentrifugal analysis*. Wiley.
23. Tersteeg, S., Mrozowich, T., Henrickson, A., Demeler, B. and Patel, T.R. (2022) Purification and characterization of inorganic pyrophosphatase for in vitro RNA transcription. *Biochemistry and Cell Biology*, **100**, 425-436.
24. D'Souza, M.H., Mrozowich, T., Badmalia, M.D., Geeraert, M., Frederickson, A., Henrickson, A., Demeler, B., Wolfinger, Michael T. and Patel, Trushar R. (2022) Biophysical characterisation of human LincRNA-p21 sense and antisense Alu inverted repeats. *Nucleic acids research*, **50**, 5881-5898.
25. Mrozowich, T., Henrickson, A., Demeler, B. and Patel, T.R. (2020) Nanoscale Structure Determination of Murray Valley Encephalitis and Powassan Virus Non-Coding RNAs. *Viruses*, **12**, 190.
26. Nelson, C.R., Mrozowich, T., Park, S.M., D'souza, S., Henrickson, A., Vigar, J.R.J., Wieden, H.-J., Owens, R.J., Demeler, B. and Patel, T.R. (2021) Human DDX17 Unwinds Rift Valley Fever Virus Non-Coding RNAs. *Int J Mol Sci*, **22**, 54.
27. Mrozowich, T., Park, S.M., Waldl, M., Henrickson, A., Tersteeg, S., Nelson, C.R., Deklerk, A., Demeler, B., Hofacker, I.L., Wolfinger, M.T. *et al.* (2022) Investigating RNA-RNA interactions through computational and biophysical analysis. *bioRxiv*, 2022.2002.2001.478553.
28. Petermann, M.L. and Hamilton, M.G. (1957) THE PURIFICATION AND PROPERTIES OF CYTOPLASMIC RIBONUCLEOPROTEIN FROM RAT LIVER. *Journal of Biological Chemistry*, **224**, 725-736.
29. Horne, C.R., Venugopal, H., Panjikar, S., Wood, D.M., Henrickson, A., Brookes, E., North, R.A., Murphy, J.M., Friemann, R., Griffin, M.D.W. *et al.* (2021) Mechanism of NanR gene repression and allosteric induction of bacterial sialic acid metabolism. *Nature Communications*, **12**, 1988.
30. Obi, J.O., Lubula, M.Y., Cornilescu, G., Henrickson, A., McGuire, K., Evans, C.M., Phillips, M., Boyson, S.P., Demeler, B., Markley, J.L. *et al.* (2020) The BRPF1 bromodomain is a molecular reader of di-acetyllysine. *Current Research in Structural Biology*, **2**, 104-115.
31. Henrickson, A., Kulkarni, J.A., Zaifman, J., Gorbet, G.E., Cullis, P.R. and Demeler, B. (2021) Density Matching Multi-wavelength Analytical Ultracentrifugation to Measure Drug Loading of Lipid Nanoparticle Formulations. *ACS Nano*, **15**, 5068-5076.
32. Soares, B.S., Rocha, S.L.G., Bastos, V.A., Lima, D.B., Carvalho, P.C., Gozzo, F.C., Demeler, B., Williams, T.L., Arnold, J., Henrickson, A. *et al.* (2022) Molecular Architecture of the Antiophidic Protein DM64 and its Binding Specificity to Myotoxin II From Bothrops asper Venom. *Front Mol Biosci*, **8**.
33. Sonn-Segev, A., Belacic, K., Bodrug, T., Young, G., VanderLinden, R.T., Schulman, B.A., Schimpf, J., Friedrich, T., Dip, P.V., Schwartz, T.U. *et al.* (2020) Quantifying the heterogeneity of macromolecular machines by mass photometry. *Nature Communications*, **11**, 1772.
34. Tanford, C. (1961) Physical Chemistry of Macromolecules John Wiley and Sons. *Inc., New York, London*, 254-269.
35. Williams, J., Van Holde, K.E., Baldwin, R.L. and Fujita, H. (1958) The theory of sedimentation analysis. *Chem Rev*, **58**, 715-744.

36. Schuster, T.M. and Toedt, J.M. (1996) New revolutions in the evolution of analytical ultracentrifugation. *Current Opinion in Structural Biology*, **6**, 650-658.
37. Sergeev, Y.V., Dolinska, M.B. and Wingfield, P.T. (2014) Thermodynamic analysis of weak protein interactions using sedimentation equilibrium. *Current protocols in protein science*, **77**, 20.13. 21-20.13. 15.
38. Chun, T., Pattem, J., Gillis, R.B., Dinu, V.T., Yakubov, G.E., Corfield, A.P. and Harding, S.E. (2023) Self-association of the glycopeptide antibiotic teicoplanin A2 in aqueous solution studied by molecular hydrodynamics. *Scientific reports*, **13**, 1969.
39. Cutajar, M., Stockman, R.A., Braovac, S., Steindal, C.C., Zisi, A. and Harding, S.E. (2022) Comparative Hydrodynamic Study on Non-Aqueous Soluble Archaeological Wood Consolidants: Butvar B-98 and PDMS-OH Siloxanes. *Molecules*, **27**.
40. Spaulding, Z., Thevarajan, I., Schrag, L.G., Zubcevic, L., Zolkiewska, A. and Zolkiewski, M. (2022) Human mitochondrial AAA+ ATPase SKD3/CLPB assembles into nucleotide-stabilized dodecamers. *Biochem Biophys Res Commun*, **602**, 21-26.
41. Zemb, T., Rosenberg, R., Marčelja, S., Haffke, D., Dufrêche, J.F., Kunz, W., Horinek, D. and Cölfen, H. (2021) Phase separation of binary mixtures induced by soft centrifugal fields. *Phys Chem Chem Phys*, **23**, 8261-8272.
42. Fedorov, D., Batys, P., Hayes, D.B., Sammalkorpi, M. and Linder, M.B. (2020) Analyzing the weak dimerization of a cellulose binding module by analytical ultracentrifugation. *Int J Biol Macromol*, **163**, 1995-2004.
43. Freyer, M.W. and Lewis, E.A. (2008) Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions. *Methods Cell Biol*, **84**, 79-113.
44. Mitra, S. (2009), *Methods in Enzymology*. Academic Press, Vol. 469, pp. 209-236.
45. Mitra, S. and Demeler, B. (2020) Probing RNA-Protein Interactions and RNA Compaction by Sedimentation Velocity Analytical Ultracentrifugation. *Methods in molecular biology (Clifton, N.J.)*, **2113**, 281-317.
46. Somarowthu, S., Legiewicz, M., Chillón, I., Marcia, M., Liu, F. and Pyle, A.M. (2015) HOTAIR forms an intricate and modular secondary structure. *Mol Cell*, **58**, 353-361.
47. Zhang, J., Pearson, J.Z., Gorbet, G.E., Cölfen, H., Germann, M.W., Brinton, M.A. and Demeler, B. (2017) Spectral and Hydrodynamic Analysis of West Nile Virus RNA–Protein Interactions by Multiwavelength Sedimentation Velocity in the Analytical Ultracentrifuge. *Analytical Chemistry*, **89**, 862-870.
48. Misra, V.K. and Draper, D.E. (2002) The linkage between magnesium binding and RNA folding. *J Mol Biol*, **317**, 507-521.

Chapter 6. Nanoscale Structure Determination of Murray Valley Encephalitis and Powassan Virus Non-coding RNAs

6.1 Foreword

This chapter consists of a manuscript wrote in collaboration with Amy Henrickson, Dr. Borries Demeler and Dr. Trushar R Patel, published in *Viruses,* Volume 12, issue 2 on February 8, 2020. https://doi.org/10.3390/v12020190. This publication focuses on the characterization of the 5' and 3' non-coding terminal regions of two flaviviruses, Murray Valley Encephalitis and Powassan Virus through computational structure prediction and small angle x-ray scattering. Amy Henrickson helped with the AUC experiments and data analysis. This publication is reproduced with permission from MDPI and the Creative Commons CC by license and re-formatted to fit thesis formatting.

6.2. Author List and Affiliations

Tyler Mrozowich [1], Amy Henrickson[1], Borries Demeler[1,2,3] and Trushar R. Patel[1,3,4,5]

1. Department of Chemistry and Biochemistry, Alberta RNA Research and Training Institute, University of Lethbridge, 4401 University Drive, Lethbridge, AB T1K 3M4, Canada
2. Department of Chemistry And Biochemistry, University of Montana, Missoula, MT 59812, USA
3. NorthWest Biophysics Consortium, University of Lethbridge, University of Lethbridge, 4401 University Drive, Lethbridge, AB T1K 3M4, Canada
4. Department of Microbiology, Immunology and Infectious Disease, Cumming School of Medicine, University of Calgary, Calgary, AB T2N 1N4, Canada
5. Li Ka Shing Institute of Virology and Discovery Lab, University of Alberta, Edmonton, AB T6G 2E1, Canada

6.3 Abstract

Viral infections are responsible for numerous deaths worldwide. Flaviviruses, which contain RNA as their genetic material, are one of the most pathogenic families of viruses. There is an increasing amount of evidence suggesting that their 5' and 3' non-coding terminal regions are critical for their survival. Information on their structural features is essential to gain detailed insights into their functions and interactions with host proteins. In this study, the 5' and 3' terminal regions of Murray Valley Encephalitis and Powassan virus were examined using biophysical and computational modeling methods. First, we used size

exclusion chromatography and analytical ultracentrifuge methods to investigate the purity of in-vitro transcribed RNAs. Next, we employed small-angle X-ray scattering techniques to study solution conformation and low-resolution structures of these RNAs, which suggest that the 3' terminal regions are highly extended as compared to the 5' terminal regions for both viruses. Using computational modeling tools, we reconstructed 3-dimensional structures of each RNA fragment and compared them with derived small-angle X-ray scattering low-resolution structures. This approach allowed us to reinforce that the 5' terminal regions adopt more dynamic structures compared to the mainly double-stranded structures of the 3' terminal regions.

6.4 Introduction

Family *Flaviviridae* consists of small, positive-sense single-stranded RNA viruses that replicate within host cells of arthropods and/or vertebrates. Flaviviruses include deadly viruses such as Murray Valley Encephalitis virus (MVEV), Powassan virus (PowV), Japanese encephalitis virus, Dengue virus, Zika virus, West Nile virus and Yellow Fever virus. The World Health Organization and the Centers for Disease Control both cite flaviviruses as a global health threat owing to the ease of transmission by mosquitoes, and the lack of efficient therapeutic or immunoprophylactic strategies (1). There is a critical need for therapeutics given the magnitude and severity of disease from this class of viruses; however, the limited understanding of viral replication and their complex interactions with the host cellular proteins through terminal region interactions hinders therapeutic development (2-4).

MVEV is a member of the Japanese encephalitis serological complex of flaviviruses and was first isolated in 1951 during the initial outbreak in Australia (5). It is believed to be contained at the top end of the Northern Territory and the North of Western Australia by a cycle involving mosquitoes (*Culex annulirostris*) and birds (6,7). The last major outbreak was in 1974 with 58 cases reported, 20% of which resulted in death (8). There have been increasing instances of MVEV, including nine cases in 2011, three of which resulted in death

(9,10). PowV is the only member of the tick-borne encephalitis serogroup that is currently present in North America (11). There has been a drastic increase in PowV over the last 18 years as compared to the previous 40 years with an increase of 671% infections in humans (12). It was first identified in Powassan, Ontario in 1958 (13) and has become endemic in the upper Midwest and the Northeastern United States, but there have also been cases reported in eastern Russia (14). The increase in PowV incidences over the last decade highlights the importance of conducting further research to understand the structure-function relationships of the viral genome to ultimately develop therapeutic approaches to combat such viral infections.

Flaviviral genomes consist of a single-stranded RNA molecule approximately 10-11k nucleotides long, depending on species. The RNA genome contains one single open reading frame, which is flanked by terminal regions (1). The ORF codes for a single polypeptide that is cleaved by a combination of viral and host proteases to produce three structural proteins and seven non-structural proteins (15). The flanking structural TRs are highly conserved across all flaviviruses and appear as complex folding sequences essential for viral replication (1,16-18) (see Supplementary Figure S6.1). It has been shown that these TRs, which are structurally conserved and intolerable to mutations/deletions, interact with the specific host proteins required for replication (19-21). The important role of these TRs is also highlighted by their ability to cyclize, where the 3’ TR interacts with the 5’ TR (22,23). This cyclization of the viral genome is essential for the correct positioning of the NS5 RNA dependent RNA polymerase produced by the viral genome (24). NS5 is required in order to produce the (-) sense RNA that is used as the replication template for further production of the viral (+) sense RNA genome (25). Flaviviral TRs also interact and are recognized by a variety of host proteins (22,26,27), including the host innate immune sensor; 2’ 5’ oligoadenylate synthetase (OAS)-family proteins and RNA helicases (DDX3X, DDX5, and DDX6) (4,20,26-29). To further understand viral replication, it is essential to study not only the complex secondary and

tertiary structures of viral TRs, but also how their structure influences their interactions with host proteins and cyclization events. For example, previous studies on the interactions of West Nile virus and adenoviral RNA with host innate immune system proteins provided insights into the RNA-protein interaction as well as mutation dependent structural changes on adenoviral RNA (22,26,30-32).

Although several attempts have been made to study RNA structures in general (33,34), detailed insight into the structural features of multiple flaviviral strains is lacking. Such insights could provide a platform to develop novel antiviral therapies as well as further our knowledge about viral replication events. Using a combination of biophysical techniques and computational calculations, we constructed 3-dimensional structures for the 5′ and 3′ TRs of both MVEV and PowV. This study provides insight into the structural organization of flaviviral TRs and provides a framework for the characterization of other flaviviral TRs, or TR segments or other non-coding RNAs.

6.5 Results

6.5.1 Purification of in vitro transcribed RNA

We purified the *in-vitro* transcribed RNAs using a Superdex 200 increase column connected to the ÄKTA FPLC unit. Figure 6.2A presents an elution profile with all four RNA constructs overlayed, where peaks at ~8 mL represent elution of template plasmids used to in-vitro transcribe these RNAs. The RNAs of interest eluted at ~13 to 14.5 mL and the peaks at ~11 mL suggest the presence of an oligomeric assembly of RNAs. After SEC, we pooled the monodispersed fractions (between ~13 to 14.5 mL) for each RNA and analyzed them using Urea-PAGE (Figure 6.2B). As demonstrated in Figure 6.2B, RNAs migrate similarly and are highly pure, devoid of any aggregation or degradation, except minor degradation of MVEV 5′ TR. These fractions were stored at 4° C until further experiments were carried out.

6.5.2 Homogeneity studies of RNA

Analytical ultracentrifugation is a versatile technique to study the purity of biomolecules in solution (35). In an AUC experiment, biomolecules are subjected to a high centrifugal force (up to 250,000 × g) to separate them on the basis of their size, anisotropy and density. The separation of biomolecules is monitored by means of an optical system. To investigate the purity of all four RNAs, we performed SV experiments at concentrations ranging between 0.5-0.7 μM and processed the data using UltraScan (36) as described in the Materials and Methods section. Figure 6.2 (C) presents the sedimentation coefficient distribution for MVEV 5′ and 3′ as well as PowV 5′ and 3′ RNA fragments. The SV analysis suggests that all four RNAs are mainly monodisperse with sedimentation coefficient values of 4.27 S for MVEV 5′ TR, 4.30 S for MVEV 3′TR, 4.49 S for PowV 5′ TR and 4.53 S for PowV 3′ TR ($S=10^{-13}$ seconds, S values corrected to 20 ºC for water), as summarized in Table 6.1. Note that the peaks at ~ 5.5 S suggest that all four RNAs form dimeric or higher-order conformations in solution. The AUC data also suggest that PowV TRs with a Mw of ~38 kDa have a slightly higher sedimentation coefficient compared to the MVEV TRs with Mw of ~32

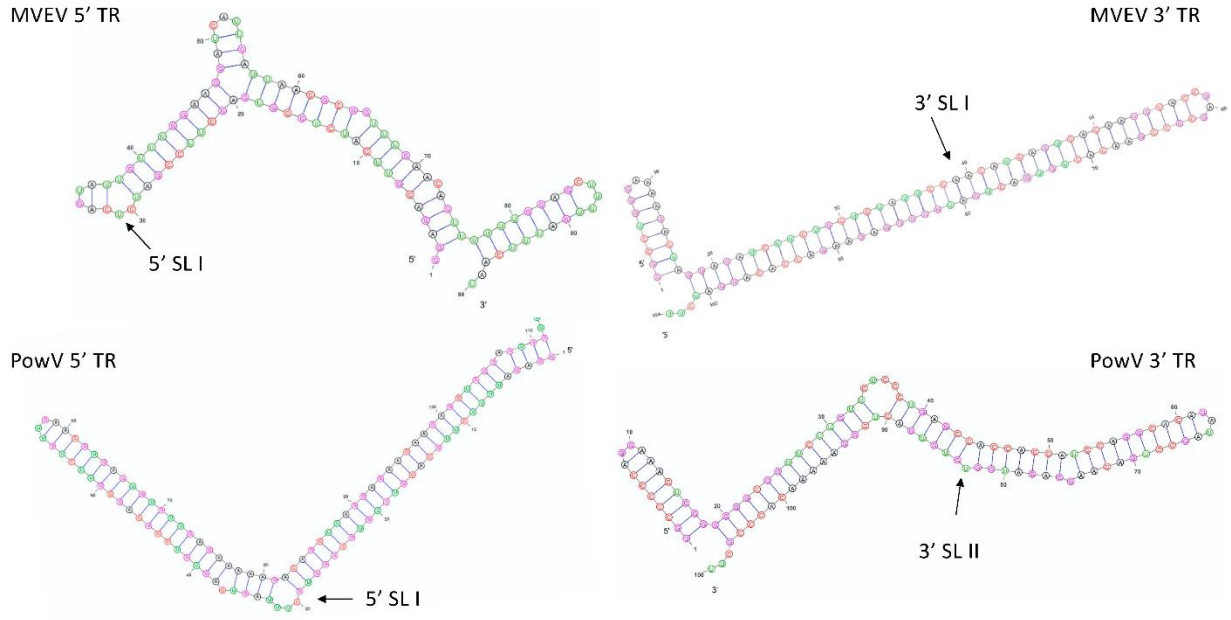


Figure 6.1. Lowest energy predicted secondary structures of MVEV, and PowV non-coding RNA regions visualized using VARNA. The next three lowest energy structures are presented in Supplementary Figure S6.2. The arrows represent stem-loop I (SLI) location in 5' and 3' TRs.

kDa (Table 6.1). Overall, these experiments indicate that all four RNAs are of suitable purity to perform HPLC-SAXS to determine the solution structure.

6.5.3 Low-resolution structural studies of RNAs

Small-angle X-ray scattering allows for structural determination of biomolecules and their complexes under physiological conditions, albeit at low-resolution. Recent improvements

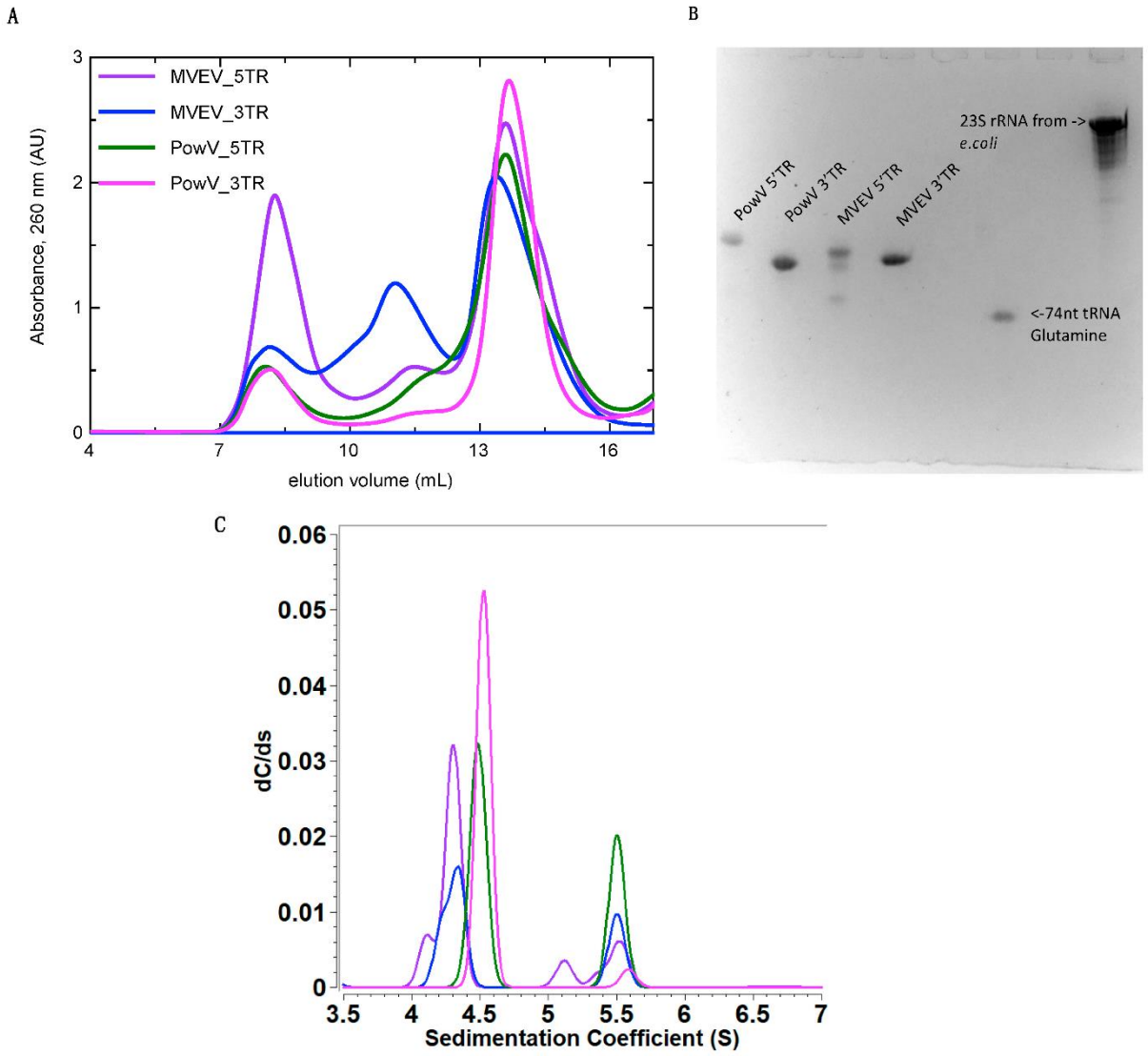


Figure 6.2. Purification and characterisation of *in-vitro* transcribed 5' and 3' TRs of MVEV and PowV RNA. **(A)** SEC elution profile of purified RNAs where the x-axis represents elution volume and the y-axis represents absorbance at 260 nm; **(B)** Urea-PAGE demonstrating the purity of the individual pooled RNA fractions from SEC peaks (~13 to 14.5 mL). For comparison purposes, we have included 74 nts tRNA and 23S rRNA from *E. coli*. **(C)** Sedimentation coefficient distribution profiles of MVEV and PowV non-coding TRs obtained from SV-AUC. The SV peaks at ~4.5 S for each RNA represent monomeric fractions. The sedimentation coefficient values are corrected to standard solvent conditions (20˚ C in water).

in instrumentation employ an HPLC unit connected in line with SAXS detection to improve the monodispersity of samples and to resolve species of interest from aggregated and degraded products (37,38). This advancement allows us to be confident in the monodispersity of our samples, even when minor heterogeneity is present. As outlined in the Materials and Methods section, we collected HPLC-SAXS data for all four RNAs, followed by the selection of data from a monodisperse peak, buffer-subtraction, and merging of selected datasets. The merged SEC-SAXS data are presented in Figure 6.3A. Subsequently, we processed the merged data using the Guinier method (plot of (I(q)) vs. (q2)), which aids detection of purity and allows determination of the R_g (average root mean squared radius from the center of the mass for a

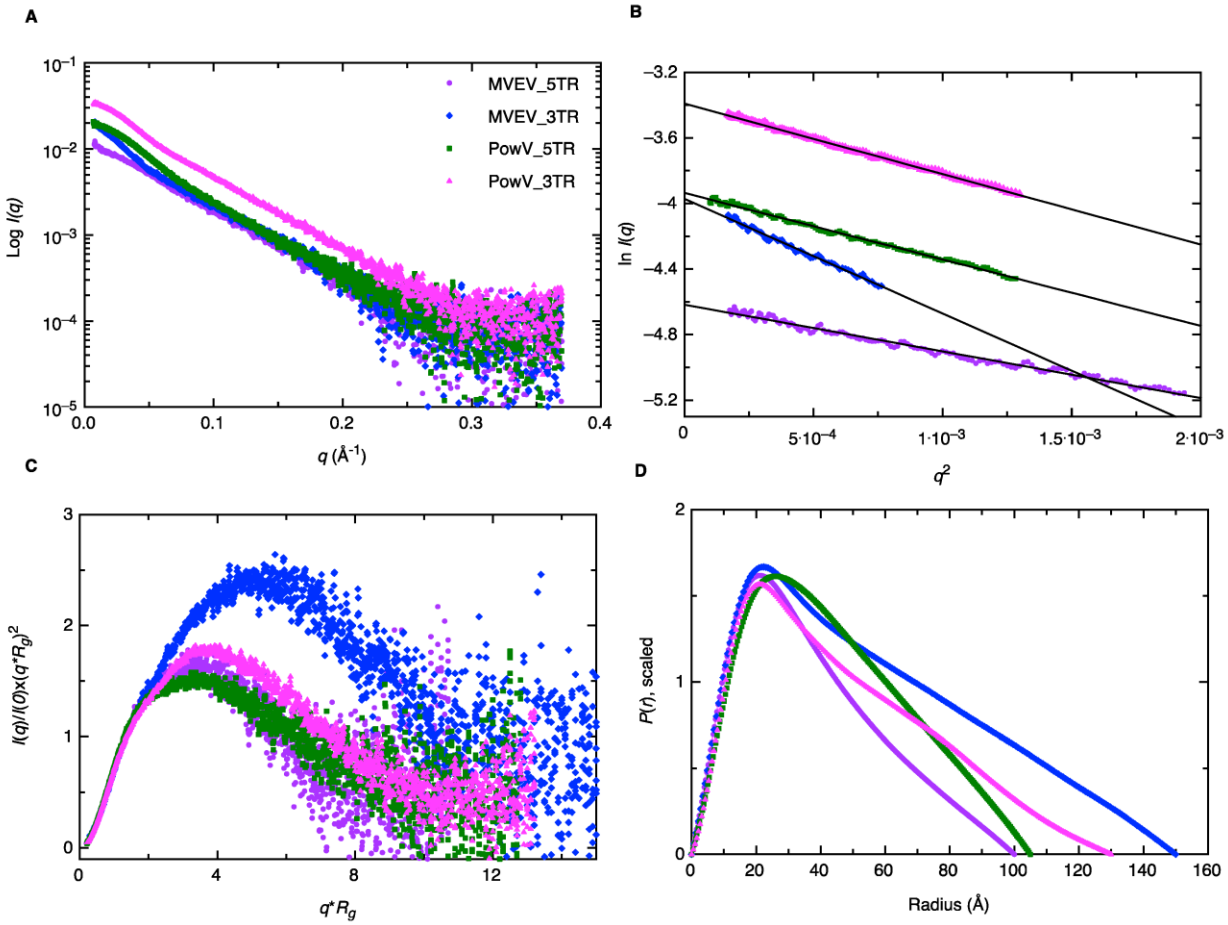


Figure 6.3. Characterization of MVEV and PowV terminal regions using SAXS. (A) A plot of scattering intensity *(Log I(q))* versus scattering angle *($q = 4\pi sin\theta/\lambda$)* representing merged SAXS data for MVEV and PowV. **(B)** Guinier plots (plot of ln(I(q)) versus q^2) representing the homogeneity of samples and allowing determination of R_g based on the low-angle region data. **(C)** Dimensionless Kratky plots *($I(q)/I(0)*(q*R_g)^2$ vs $q*R_g$)* for all four RNA samples demonstrating their extended structures. **(D)** Pair-distance distribution (P(r)) plots for all four RNA samples representing their maximal particle dimensions and allowing the determination of R_g from the entire SAXS dataset.

biomolecule) from the data belonging to the low-q region [45]. Figure 6.3B presents the Guinier plots for 5′ and 3′ TRs of MVEV and PowV, where the linearity for low-q data demonstrates that all four RNAs are monodispersed and devoid of any aggregation. Based on the Guinier analysis, we obtained R_g values of 29.01 ± 0.22 Å, 45.72 ± 0.48 Å, 34.65 ± 0.19 Å and 35.84 ± 0.13 Å for MVEV 5′ TR, MVEV 3′ TR, PowV 5′ TR and PowV 3′ TR respectively (see Table 6.1 for more details). Once we confirmed the monodispersity, we processed the SAXS scattering data from Figure 6.3A to obtain dimensionless Kratky plots $(I(q)/I(0)*(q*R_g)^2$ vs $q*R_g$) that allow detection of the folding state of biomolecules (38,39). For example, globular-shaped biomolecules in solution are observed with a well-defined maximum value of 1.1 at $q*R_g = 1.73$ (40). The dimensionless Kratky plots for 5′ and 3′ TR of MVEV and PowV under investigation demonstrate that all the samples are well folded, and extended in solution (Figure 6.3C).

Table 6.1. Biophysical parameters of MVEV and PowV non-coding RNA.

Sample	MVEV 5′TR	MVEV 3′TR	PowV 5′TR	PowV 3′TR
M_w (kDa, sequence)	31.38	31.57	36.92	37.80
Sedimentation coefficient (S)$^{\nabla}$	4.27 (4.08, 4.46)	4.30 (4.20, 4.41)	4.50 (4.42, 4.55)	4.53 (4.52, 4.53)
I(0)$^{\#}$	0.0098 ± 4.7 x 10^5	0.019 ± 1.4 x 10^5	0.019 ± 6.7 x 10^5	0.034 ± 8.2 x 10^5
q.Rg range	0.38 – 1.26	0.59 – 1.26	0.34 – 1.26	0.47 – 1.29
R_g (Å)$^{\#}$	29.01 ± 0.22	45.72 ± 0.48	34.65 ± 0.19	35.84 ± 0.13
I(0)$^{\#\Delta}$	0.098 ± 3.8 x 10^5	0.018 ± 8.1 x 10^5	0.019 ± 4.2 x 10^5	0.034 ± 8.9 x 10^5
R_g (Å)$^{\Delta}$	30.16 ± 0.11	46.16 ± 0.18	34.48 ± 0.06	38.39 ± 0.08
D_{max} (Å) $^{\Delta}$	100	150	105	130
χ^{2*}	~0.78	~0.83	~0.78	~0.80
NSD*	0.71 ± 0.01	0.80 ± 0.02	0.85 ± 0.02	0.72 ± 0.01

The Mw values were calculated using nucleotide sequences. ∇- determined using SV-AUC analysis and UltraScan-III package. Sedimentation coefficients obtained following genetic algorithm-Monte Carlo analysis. The measured value represents the mean. Bracketed values represent a 95% confidence interval from Monte Carlo analysis. # - obtained from Guinier analysis. Δ - determined using P(r) analysis using the GNOM program. * - values derived from DAMMIN and DAMAVER analysis.

Next, using program GNOM (41), we performed an indirect Fourier transformation to convert the reciprocal-space information (ln(I(q)) vs. (q), Figure 6.3A) into the real space electron pair-distance distribution function (P(r), Figure 6.3D) to obtain R_g and D_{max} (maximum particle dimension) for all four RNAs. It is important to note that unlike Guinier analysis, which is restricted to the data in the low-q region, the P(r) analysis utilizes a wider-range of the dataset and aids in a reliable determination of R_g and D_{max}. As outlined in Table 6.1, based on P(r) analysis, we obtained the D_{max} of 100 Å, 150 Å, 105 Å, and 130 Å for MVEV 5′ TR, MVEV 3′ TR, PowV 5′ TR, and PowV 3′ TR respectively. Furthermore, we obtained the R_g values of 30.16 ± 0.11 Å, 46.16 ± 0.18 Å, 34.48 ± 0.06 Å and 38.39 ± 0.08 Å for MVEV 5′ TR, MVEV 3′ TR, PowV 5′ TR and PowV 3′ TR respectively. These values are very similar to those obtained from Guinier analysis, which indicates that the data are suitable for low-resolution shape reconstruction. The shape of the P(r) plot is indicative of the solution conformation of biomolecules. For example, we would have expected a bell-shaped P(r) distribution curve with a maximum at ~D_{max}/2 (42) for a globular-shaped protein; however,

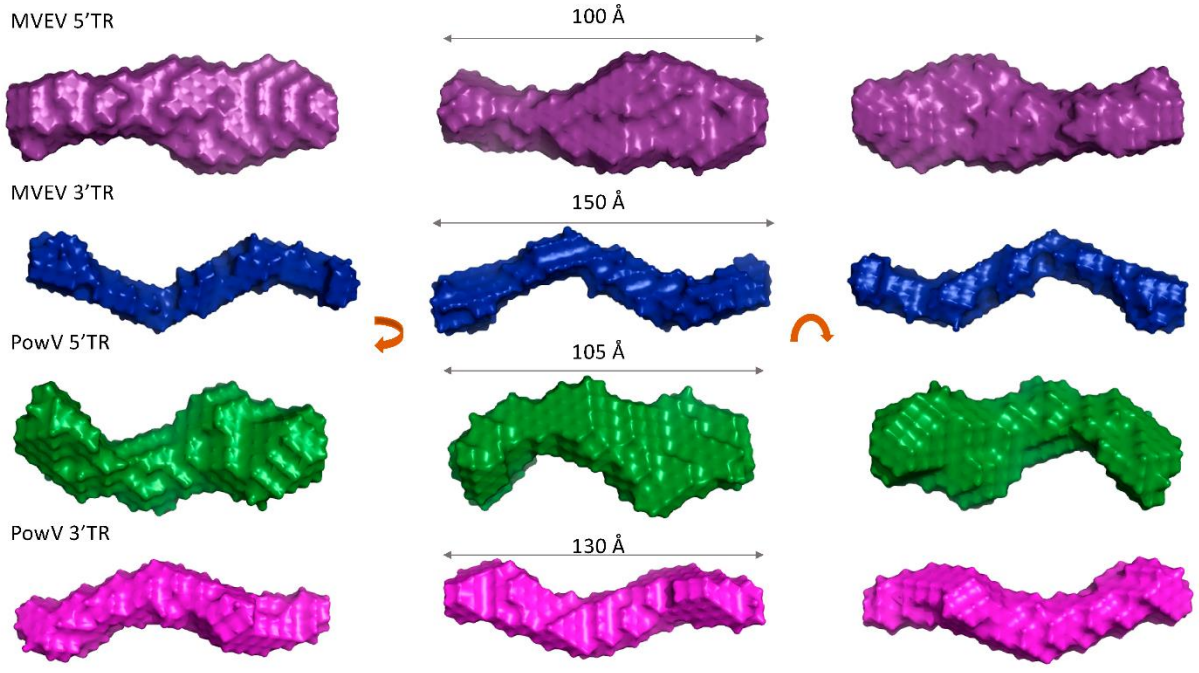


Figure 6.4. Determination of low-resolution structures using SAXS. Models indicate that all RNA molecules adopt an extended structure in solution. For each terminal region RNA, the left and right panels represent 180° rotated view on the x-axis and y-axis, respective of the view presented in the middle panel. Dimensions represent the D_{max}.

all four RNAs display skewed bell-shaped curves with extended tails that suggest their extended structures in solution as is shown in Figure 6.3D.

To obtain low-resolution structures for each RNA, we employed DAMMIN (43) that utilizes a simulated-annealing protocol and allows the incorporation of P(r) data (i.e., D_{max} and R_g as constraints). We calculated a total of 12 models for 5′ and 3′ TRs of MVEV and PowV and noted that individual models had an excellent agreement between the experimentally obtained scattering data and the calculated scattering data. The χ^2 values in each case are ~0.8, which represents an agreement between the experimentally collected scattering data and the low-resolution model derived scattering data (see Table 6.1). Next, we employed

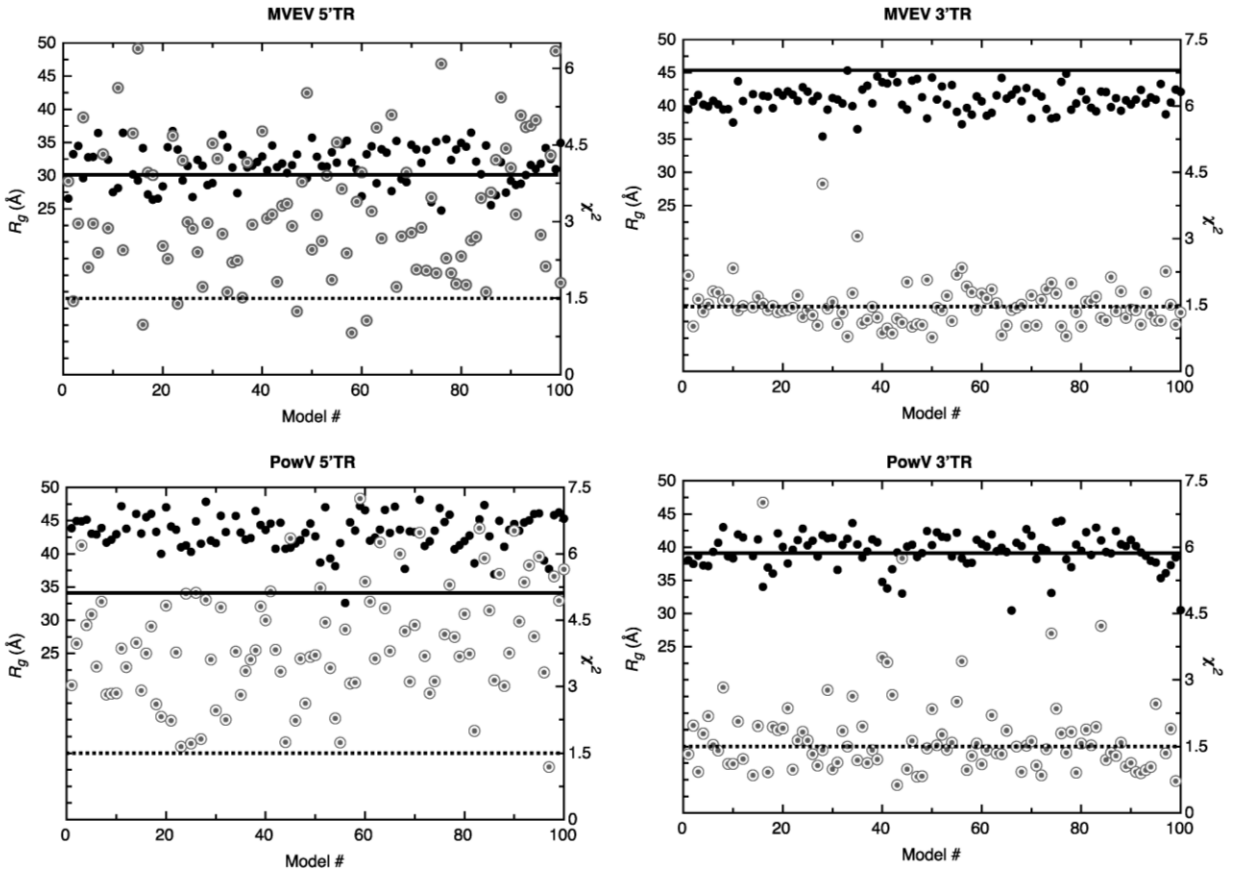


Figure 6.5. Screening of high-resolution structures calculated using MC-Sym for MVEV 5' TR, MVEV 3'TR, PowV 5'TR and PowV 3'TR. For each RNA, the x-axis represents model numbers (total 100 models), whereas the y1-axis (solid black circles) and y2-axis (grey double circles) represent Rg (Å) and χ^2 values, respectively, calculated using CRYSOL package. For each plot, the solid dark line corresponds to y1-axis and represents the value for experimentally determined R_g (Å) as presented in Table 6.1. The dotted line corresponds to y2-axis and represents a χ^2 value of 1.5. These plots indicate that MC-Sym derived structures represent a wide-range of conformations these RNAs can theoretically adapt.

program DAMAVER (44) to rotate and align all 12 models and to obtain an averaged filtered structure for each RNA, that represents structural features from individual models (Figure 6.4) (44). In each case, the goodness of the superimposition of individual models was estimated by the overlap function—the normalized spatial discrepancy (NSD). As presented in Table 6.1, the low NSD values suggest that 12 models in each case are highly similar to each other. Figure 6.4 presents the averaged filtered structures for the 5' and 3' TRs of MVEV and PowV, which indicates that, overall, these RNAs have extended structures in solution.

6.5.4 Computational modeling of RNA structures

SAXS is an outstanding method that allows the determination of solution structures of biomolecules, albeit at low-resolution. Techniques such as X-ray crystallography and nuclear magnetic resonance spectroscopy provide high-resolution structural information but have limited applications for RNA structural studies due to challenges obtaining high-quality crystals and RNA-labeling, respectively. An alternative to this approach is to employ computational

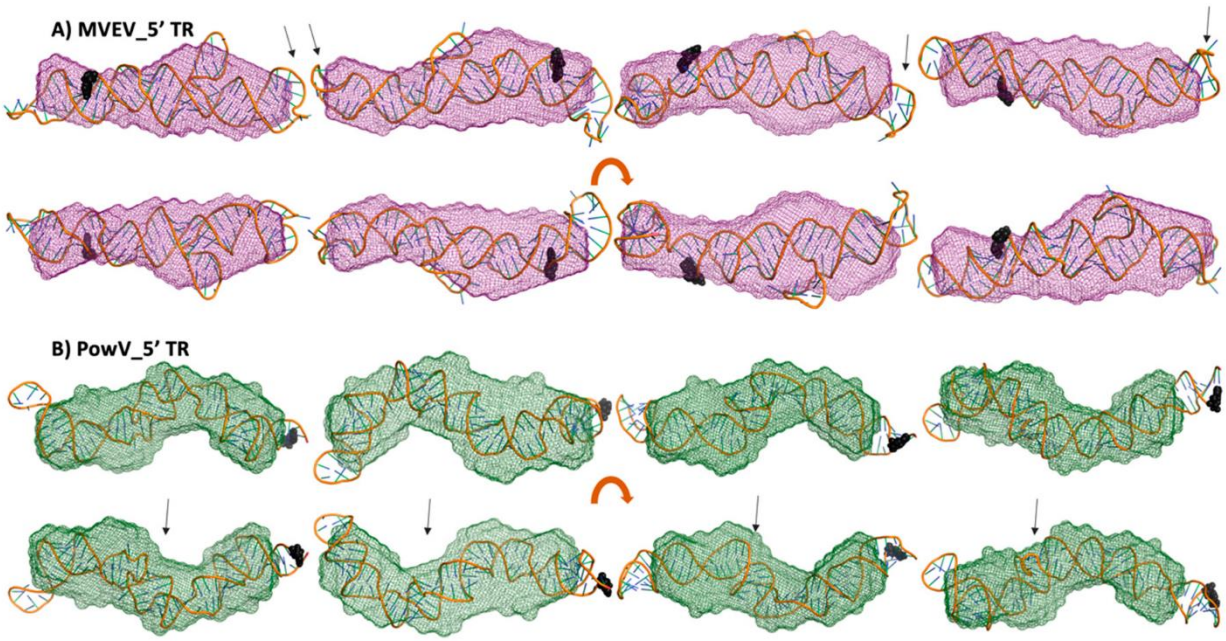


Figure 6.6. High-resolution structures calculated using MC-SYM overlaid with low-resolution SAXS models. A) MVEV 5TR **B)** and PowV 5TR. Bottom panels in both cases represent a 180o rotation along the x-axis represented in the top panels. Black spheres represent 5' terminal region on each construct. The arrows represent stem-loop I (SLI) location in 5' TRs.

modeling of RNA structures and screen those structures using experimental data. We employed the MC-Fold/MC-Sym pipeline developed by Parisien and Major (45). For each RNA, we calculated 100 structures using the MC-Sym pipeline based on the secondary structure predicted using MC-Fold (Figure 6.1) and visualized with VARNA (46). ΔG values for the lowest energy structures for MVEV 5'TR, MVEV 3'TR, PowV 5'TR, and PowV 3'TR are -86.41, -104.44, -103.07, and -103.07 kcal mol-1, respectively. Next, we used program CRYSOL (47) to calculate X-ray scattering profiles of 100 structures of each RNA (5' and 3' TRs of MVEV and PowV) and compared these data with experimentally collected X-ray scattering data to calculate χ^2 value. We also determined R_g from MC-Sym derived structures. Figure 6.5 presents the R_g (y-axis left-hand side, solid black circles) and χ^2 values (y-axis right-hand side, grey double circles) for each RNA system, which suggests that the RNA molecules under investigation can adopt a variety of conformations. For example, in the case of 5' TR of MVEV, the distribution of R_g values range from 30 ± 5 Å, whereas the experimentally determined R_g for this construct is 30.16 Å. Similarly, for PowV 3' TR, experimentally determined R_g is 38.39 Å, which largely aligns with the distribution of R_g values of MC-Sym derived structures. On

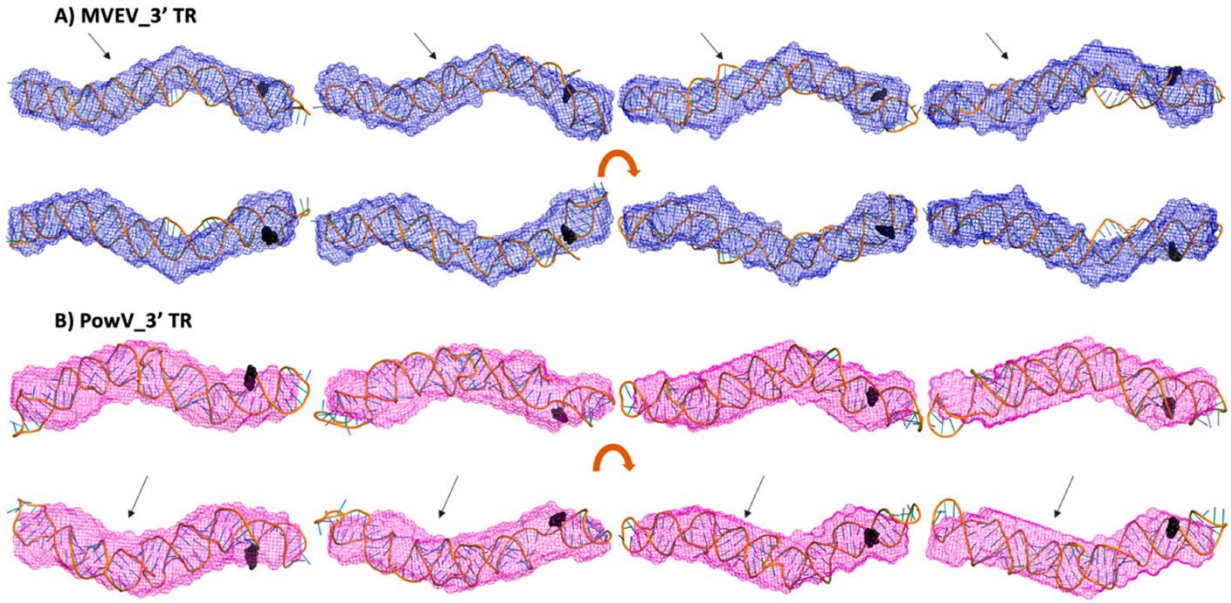


Figure 6.7. High-resolution structures calculated using MC-SYM overlaid with low-resolution SAXS models. A) MVEV 3TR **B)** and PowV 3TR. Bottom panels in both cases represent an 180o rotation along the x-axis represented in the top panels. Black spheres represent 5' terminal region on each construct. The arrows represent stem-loop I (SLI) location in 3' TRs.

the other hand, the MC-Sym derived models for MVEV 3′ TR mainly under-represent the R_g values when compared to experimental R_g of 46.16 Å. In contrast, for PowV 5′TR, Mc-Sym derived structures have a higher R_g distribution (~45Å) compared to the experimentally determined R_g (34.48 Å). Moreover, as presented in Figure 6.5, the χ^2 values for MC-Sym derived 3′ TR of MVEV and PowV structures have better distribution around 1.5 χ^2, compared to MVEV and PowV 5′ TRs.

6.5.5 Combination of Computational Modeling and Experimental Low-Resolution SAXS Structures

We selected ~10 MC-Sym derived structures with the lowest χ^2 values to further investigate if they align with the low-resolution averaged filtered models for each RNA system. Figures 6 and 7 present four of the best χ^2 fit MC-Sym derived structures for MVEV and PowV, indicating that these structures agree well with the low-resolution models based on SAXS data. Overall, MC-Sym derived structures for the 3′ TRs of MVEV and PowV (Figure 6.7) indicate that they adopt highly extended structures mainly in the double-stranded regions. In contrast, the 5′ TRs of MVEV and PowV are more dynamic with less double-stranded regions (Figure 6.6). Furthermore, these results indicate that the 5′ TR structures have a higher content of un-paired/loop regions compared to the 3′ TRs which contains mainly double-stranded regions.

6.6 Discussion

MVEV is endemic to northern Australia and has seen an increase in prevalence during the last decade. The likelihood of an outbreak is increasing given the extensive human development in affected areas. PowV is an under-studied flavivirus despite its increasing incidence and severe neurological effects. Patients infected with almost any flavivirus, including PowV and MVEV, are treated only with symptomatic support. Treatment of the viral infection is nonexistent. Further investigation into these potentially deadly viruses is

necessary to gain detailed insights into the viral life cycle in order to prepare ourselves better for potential future outbreaks.

The lowest energy secondary structure predictions (Figure 6.1) reveal that each RNA molecule varies slightly due to increasing energy, except for PowV 5’ TR and to a lesser degree MVEV 3’ TR. The structure with the lowest predicted energy of PowV 5’ TR displays a large change in secondary structure with only a small change in energy (-103.07 vs -102.43). The secondary structure information for MVEV 3’ TR indicates that a small increase in energy could result in the formation of a second stem-loop with a small energy change (-104.44 to -103.29, Supplementary Figure S6.2). This suggests that the RNA molecules can adopt different conformations in solution.

We prepared all four RNAs using in-vitro transcription with T7 polymerase and purified using SEC (Figure 6.2B), which indicates that these RNAs can form dimer/oligomers in solution. The fractions indicating a homogenous preparation were pooled together and analyzed using Urea-PAGE (Figure 6.2B), where the three RNAs displayed a single band, and MVEV 5’TR showed a slight degradation. We used an RNA that was smaller than our constructs (74nt tRNA), and another, which was larger, (23S rRNA) to represent a relative difference in RNA size (Figure 6.2B), (Table 6.1). We further employed SV experiments using analytical ultracentrifugation to determine the composition and anisotropy of biomolecules in solution (35,48,49). Figure 6.2 presents the sedimentation coefficient distribution of SEC-purified RNA preparation. As evident by the peaks between 5S and 6S, the RNAs can form higher-order oligomers to a small degree despite removing many of them by means of SEC purification. This is consistent with the HPLC data from SAXS, which also shows a minor amount of oligomerization for each RNA. An explanation for these minor oligomerization peaks could be the RNA adopting different conformations, as seen in the predicted secondary structure(s) of PowV 5’ TR (Supplementary Figure S6.2) where a small energy change caused the RNA to adopt a different conformation. This prediction could explain the increased peak that is evident

at 5.5 S (Figure 6.2C) for both PowV 5′ TR and MVEV 3′ TR. Additionally, the peak at ~ 4 S for MVEV 5′ TR indicates minor degradation, which was also seen in the Urea-PAGE (Figure 6.2C and 6.2B). More importantly, the differences in the RNA size and shape are indicated by their sedimentation coefficient values (Table 6.1). Both the TRs of MVEV have a Mw of ~ 32 kDa and produce very similar S values (4.27 S and 4.3 S, for 5′ and 3′ TR, respectively). Furthermore, the S values for 5′ and 3′ TRs of PowV are 4.50 S and 4.53 S. This similarity correlates with the similarities of their Mw of ~ 37 kDa (Table 6.1).

SAXS is an excellent complementary structural-biophysical method, which enables solution structural studies of RNA, proteins, and their complexes, albeit at low-resolution (33,37,39,50-53). Solution X-ray scattering is employed for biological systems where obtaining high-quality crystals or labeling of biomolecules presents challenges (22,30,31,50,52,53). In this study, we employed an HPLC-SAXS set-up to collect scattering data from a preparation free of aggregation or degradation. The monodispersed preparation was confirmed by Guinier analysis and shows excellent linearity of fit in the low-q region (Figure 6.3B). Furthermore, Guinier analysis also provided R_g of all four RNAs (based on the low-q region), which are highly similar to those calculated by means of P(r) analysis (Figure 6.3D, Table 6.1). We also performed dimensionless Kratky analysis (Figure 6.3C) that demonstrates each RNA adopts an elongated structure, with a low amount of flexibility. The P(r) distribution reveals a quick increase to the highest point and then a gradual decrease where the electron-pair-distance approaches zero. A globular molecule would result in a gaussian-like distribution, which is not evident in any of the RNA and suggests their elongated nature. The most extended molecule is MVEV 3′ TR resulting in a D_{max} of 150 Å, even though its molecular weight is 31.57 kDa (Table 6.1), which is lower than both 5′ and 3′ TR of PowV (Table 6.1). Low-resolution structure modeling (Figure 6.4) confirms that MVEV 3′ TR adopts a much thinner and elongated structure, which suggests that this RNA is almost entirely helical. This is also evident for PowV 3′ TR, which also displays an extended conformation.

This elongation is consistent with the predicted secondary structure of flaviviral 3′ terminal regions where it is suggested the last ~100 nucleotides almost entirely base pair into a large system (22,54-56). Low-resolution structures of 5′ TRs of MVEV and PowV are less elongated when compared to the 3′ TRs; however, their low-resolution structures appear to be consistent with the predicted secondary structures of flaviviral 5′ TRs. This indicates that the 5′ TRs adopt more dynamic structures (26,54-56). Overall, MVEV and PowV TR structures are consistent with previously published low-resolution structures of flaviviral RNA such as the Zika and West Nile viruses (22,26,57). Although the flaviviruses possess a methylated type 1 cap (me7-GpppA-me2) on the 5′ terminus, our *in vitro* transcribed RNAs are not methylated, and we anticipate that an addition of an extra guanine residue through a non-standard 5′ to 5′ triphosphate linkage would not affect the low-resolution structure determination. Furthermore, due to T7 requiring a guanine residue to start transcription, our constructs actually contain a G nucleotide on the 5′ end similar to a type 1 cap.

A strength of SAXS is its ability to be combined with high-resolution structures/homology models of individual domains, as well as with computational studies to complement predicted structures (22,30,39,50,53). RNA structures of high resolution are hard to determine through crystallization and labeling studies mentioned above; therefore, an attractive alternative to calculating high-resolution structures is computational modeling. To this end, we employed the MC-Fold/MC-SYM pipeline to reconstruct all-atom structures and screen those using experimentally collected X-ray scattering data to identify structures that are likely adopted by these RNAs. For each TR structure, we used CRYSOL (47) to calculate R_g as well as back-calculated X-ray scattering data. We aligned this data with experimentally collected scattering data (Figure 6.5). This analysis demonstrates that MVEV 3′ TR and PowV 3′ TR exhibit a strong correlation between the experimentally determined R_g and the calculated R_g from MC-Sym-derived structures. For both of these RNAs the χ^2 values indicate agreement between scattering data that were experimentally collected and calculated

from MC-Sym-derived structures, where the χ^2 values were close to ~1.5 for most structures (Figure 6.5). On the other hand, the χ^2 values for MC-Sym-derived structures for 5′ TRs of MVEV and PowV have a much larger distribution (from ~1.5 to ~6). A similar trend was also observed for R_g for both 5′ TRs where the MC-Sym derived structures displayed a wider distribution of R_g, suggesting that these RNAs could adopt multiple conformations. The correlation between R_g and χ^2 was far more pronounced in the more elongated TR RNAs - PowV 3′ TR and MVEV 3′ TR (Figures 6.4 and 6.5). Finally, we selected the ~10 MC-Sym-derived structures, which had the lowest χ^2 values when compared to experimental data and aligned them into the low-resolution SAXS envelopes. Selected groups of structures are presented in Figures 6.6 and 6.7, which indicate that the computationally calculated structures agree well with the experimental data, especially for the 3′ TRs. For the 5′ TRs, we observed dynamic structures; however, they maintain a similar conformation in solution overall. Furthermore, the bulge regions in both the 5′ TRs also agree well with their secondary structures, which suggests that these secondary structures could be conserved and do not undergo conformational changes. An added benefit of these computationally derived structures is the addition of directionality of the RNAs (the black spheres represent the 5′ end of each molecule in Figures 6.6 and 6.7), which is otherwise very challenging to decipher unless the molecule of interest is altered (58).

6.7 Conclusions

In this study, we have in-vitro transcribed MVEV and PowV terminal RNA regions and performed their native purification, studied their homogeneity using analytical ultracentrifugation, determined their low-resolution structures, and combined computational modeling to obtain high-resolution structural models. Detailed information on the molecular biology of flavivirus and their life cycle is already available (e.g. (59)). However, there are limited biochemical and biophysical studies of flaviviral nucleic acids, their interactions with host proteins, and the functional role of flaviviral nucleic acids – host protein interactions.

Without detailed information on flaviviral nucleic acid tertiary structures, it will be challenging to obtain a complete picture of how flaviviral TRs play a part in viral replication. Recently, it was shown for the West Nile virus that the 3' stem-loop region undergoes a temperature-dependent genome cyclization, meaning that the higher temperatures in humans facilitate genome cyclization and viral replication, which does not occur in mosquitos due to their decreased body temperature (60). Potential temperature-based change in viral RNA tertiary structure is an important aspect to investigate further. We demonstrated that to gain high-resolution viral RNA structural details, a combination of SAXS and computational methods is feasible. This approach will help us understand the role of flaviviral tertiary structure in viral replication, through the establishment of a pipeline that allows for 3-dimensional visualization of RNA and RNA-protein complexes.

6.8 Materials and Methods

6.8.1 RNA Preparation and Purification

We prepared cDNA sequences under the control of a T7 RNA polymerase promoter with two additional G nucleotides on the 5' end followed by an *XbaI* restriction enzyme cut site (T^CTAGA). We designed our constructs for MVEV and PowV based on the Genebank sequences of KX229766.1and EU670438.1, respectively. All RNA constructs used in the experiments are listed as follows:

1, MVEV 5TR 1-96nt

5'GGAGACGUUCAUCUGCGUGAGCUUCCGAUCUCAGUAUUGUUUGGAAGGAUCAUUGAUUAACGC
GGUUUGAACAGUUUUUUGGAGCUUUUGAUUUCAAU3'

2, MVEV 3TR 10914-11014

5'GGCCUGGGAAAAGACUAGGAGAUCUUCUGCUCUAUUCCAACAUCAGUCACAAGGCACCGAGCG
CCGAACACUGUGACUGAUGGGGGAGAAGACCACAGGAUCUU3'

3, POWV 5TR 1-111

5’GGAGAUUUUCUUGCACGUGUGUGCGGGUGCUUUAGUCAGUGUCCGCAGCGUUCUGUUGAACG
UGAGUGUGUUGAGAAAAAGACAGCUUAGGAGAACAAGAGCUGGGAGUGGUUU3’

4, POWV 3TR 10735-10839

5’GGCCCCCAGGAAACUGGGGGGGCGGUUCUUGUUCUCCCUGAGCCACCACCAUCCAGGCACAGA
UAGCCUGACAAGGAGAUGGUGUGUGACUCGGAAAAACACCCGCUU3’

We performed *in vitro* transcription reactions using T7 RNA polymerase to prepare each RNA, followed by their purification using a Superdex 200 increase (GE Healthcare Canada inc, Mississauga, ON) on an ÄKTA pure FPLC (fast protein liquid chromatography) system (GE Healthcare) at 0.5mL/min. We collected and pooled fractions containing purified RNAs from the size exclusion chromatography (SEC) fraction collector. Pooled fractions were ethanol precipitated and resuspended in RNA buffer (50 mM Tris pH 7.5, 100 mM NaCl and 5 mM MgCl2). Next, we analyzed the pooled fractions using Urea-Poly Acrylamide Gel Electrophoresis (Urea-PAGE). We mixed 10 µL of ~500 nM RNA with 2 µL of RNA loading dye and placed them into a 1.0 cm well PAGE casting plate (Bio-Rad Laboratories (Canada), Mississauga, ON). Finally, we ran the Urea-PAGE (7.5%) at 300V at room temperature for 25 minutes in 0.5 X TBE, followed by staining with Sybr Safe (Thermofisher Scientific, Saint-Laurant, QC) and UV visualization. Prior to performing AUC and SAXS experiments, we heated RNA to 95 ºC for 5 min and cooled passively to room temperature to facilitate refolding.

6.8.2 Analytical Ultracentrifugation (AUC)

The sedimentation velocity AUC data for FPLC-purified RNA samples were collected using a Beckman Optima AUC centrifuge and an AN50-Ti rotor at 20 °C. We loaded MVEV 5’ (0.6 µM), MVEV 3’ (0.5 µM), PowV 5’ (0.6 µM) and PowV 3’ (0.58 µM) samples into Epon-2 channel centerpieces in 50 mM Tris pH 7.5, 100 mM NaCl, and 5 mM $MgCl_2$ buffer. As a first step, we centrifuged samples at 35,000 revolutions per minute and collected scans at 20-

second intervals. We used the UltraScan-III package (36) to analyse all data using Comet (San Diego Supercomputing Center, San Diego, CA) and Lonestar5 (Texas Advanced Computing Center, Austin, TX). We initially analyzed the SV-AUC data using two-dimensional spectrum analysis (2DSA) with simultaneous removal of time-invariant noise, meniscus and bottom positions fitted (61) followed by enhanced van Holde-Weischet analysis (62) and genetic algorithm refinement (63). We derived fitting statistics from the Monte Carlo analysis (64) and estimated the buffer density and viscosity corrections with UltraScan (1.0030g/cm3 and 1.0100 cP, respectively). All hydrodynamic parameters were corrected to standard conditions at 20 ºC and water.

6.8.3 Small-angle X-ray scattering (SAXS)

We utilized the B21 beamline at Diamond Light Source (Didcot, Oxfordshire, UK) to collect small-angle X-ray scattering (HPLC-SAXS) data as previously described (65). Making use of an in-line Agilent 1200 (Agilent Technologies, Stockport, UK) HPLC connected to a flow cell, each purified RNA, 50 µL of ~2.0 mg mL-1 was injected into a buffer equilibrated Shodex KW403-4F (Showa Denko America Inc., New York, NY) size exclusion column at a flow rate of 0.160 mL per minute. Each frame was exposed to the X-rays for 3 seconds. Each sample peak region was buffer subtracted and merged using Primus (66) or ScÅtter (67) as described previously (68). We analyzed the merged data using Guinier approximation to obtain the radius of gyration (R_g) and study homogeneity of samples (69). We also performed dimensionless Kratky analysis (40) to investigate if the RNA molecules of interest are folded, as reviewed earlier (39). Next, we performed the pair-distance distribution (P(r)) analysis using the program GNOM (41), which provided the R_g and the maximum particle dimension (D_{max}, the radius at which the P(r) dependence approaches zero). We used information from the P(r) plot to generate the models using DAMMIN (43), with no enforced symmetry, as described previously (53). Lastly, we averaged and filtered the resulting models to obtain a

singular representative model using the DAMAVER package (44), as described previously (31,70).

6.8.4 Atomic Structures Calculations

To generate secondary structure predictions used downstream, we used MC-Fold (45) to predict numerous low energy secondary structures for 5′ and 3′ TRs of MVEV and PowV (Figure 6.1). We selected the lowest energy structures in each case as input files for MC-Sym that enables the fragment-based reconstruction of 3-dimensional structures using known structures, as described earlier (45). Through MC-Sym, we calculated 100 all-atom structures for each RNA and minimized using protocols implemented in MC-Sym. Next, we simulated scattering profiles for each MC-Sym calculated structure with CRYSOL based on minimized structures and compared against experimental SAXS data. We subjected the minimized structures to CRYSOL to determine R_g and the goodness-of-fit parameter (χ^2) (47). Finally, we ranked the MC-Sym derived structures based on their χ^2 values and aligned them with low-resolution structures using the program SUPCOMB (71).

6.9 Supplemental Materials

5′ TR comparison (first 101 nt)

>PowV 5'TR

AGAUUUUCUUGCACGUGUGUGCGGGUGCUUUAGUCAGUGUCCGCAGCGUUCUGUUGAACGUGA
GUGUGUUGAGAAAAAGACAGCUUAGGAGAACAAGAGCU

>MVEV 5'TR

CCUGGGAAAAGACUAGGAGAUCUUCUGCUCUAUUCCAACAUCAGUCACAAGGCACCGAGCGCCGA
ACACUGUGACUGAUGGGGGAGAAGACCACAGGAUCU

>DENV 5'TR

GCUUAACGUAGUGCUGACAGUUUUUUAUUAGAGAGCAGAUCUCUGAUGAACAACCAACGGAAGAA
GACGGGAAAACCGUCUAUCAAUAUGCUGAAACGCGU

>WENV 5'TR

UGACAAACUUAGUAGUGUUUGUGAGGAUUAACAACAAUUAACACAGUGCGAGCUGUUUCUUAGC
ACGAAGAUCUCGAUGUCUAAGAAACCAGGAGGGCCCG

3’TR comparison (First 101 nt)

>PowV 3'TR

CCCCCAGGAAACUGGGGGGGCGGUUCUUGUUCUCCCUGAGCCACCACCAUCCAGGCACAGAUAG
CCUGACAAGGAGAUGGUGUGUGACUCGGAAAAACACC

>MVEV 3'TR

CCUGGGAAAAGACUAGGAGAUCUUCUGCUCUAUUCCAACAUCAGUCACAAGGCACCGAGCGCCGA
ACACUGUGACUGAUGGGGGAGAAGACCACAGGAUCU

>DENV 3'TR

UCUUGGACUAUAUGCCUUCAAUGAAGAGGUUCAGGAAGGAAGAGGAGUCGGAGGGGGCCAUUUG
GUAAACGUAGGAAGUGAAAAAGAGGCUAACUGUCAGG

>WENV 3'TR

AUGAAGACACAACUUUAGUUGAGGACACAGUACUGUAAAUACUUUAUCAAUUGUAAAUAGACAAU
AUAAGCAUGUAUAUAGGUAUGGGUUUAUAGUGGCAU

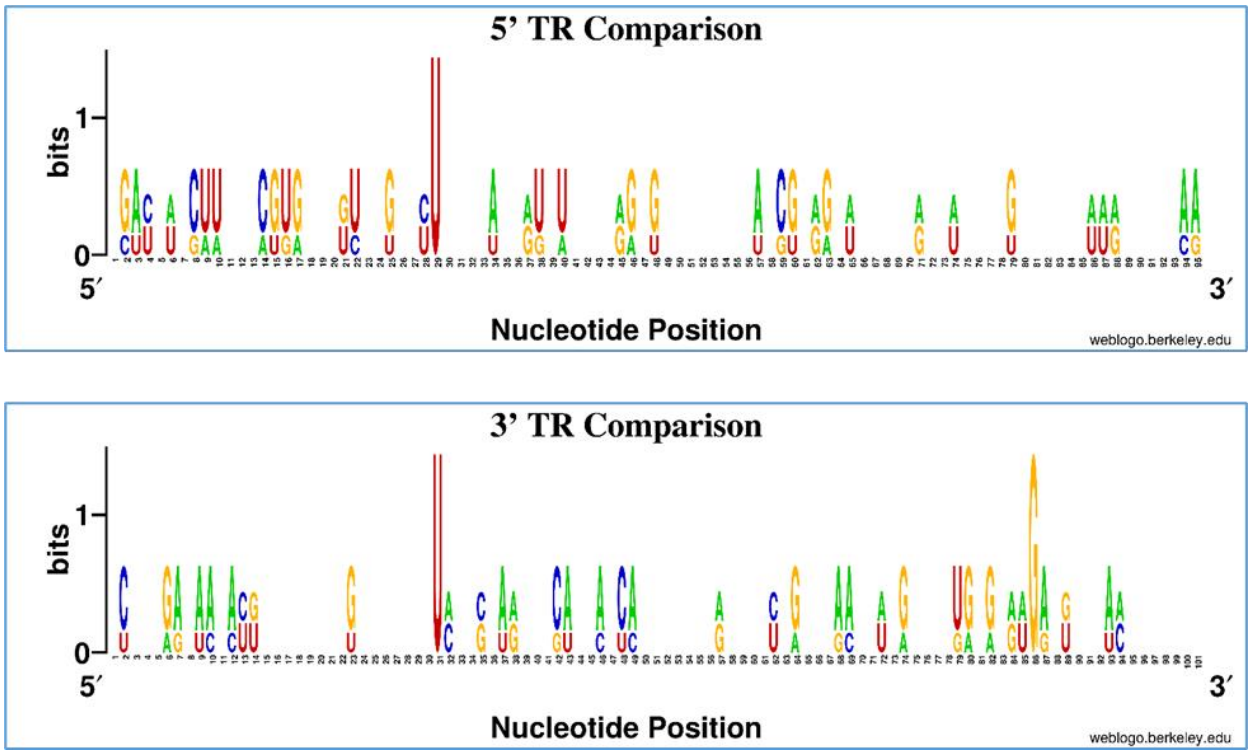


Figure S6.1. Sequence alignment of 5’ (top panel, first 101 nucleotides) and 3’ (bottom panel, last 101 nucleotides) TRs of Powassan, Murray valley encephalitis, dengue, and West Nile viruses. The sequences used for alignment are presented below. The alignment was performed using WEBLOGO (https://weblogo.berkeley.edu/logo.cgi).

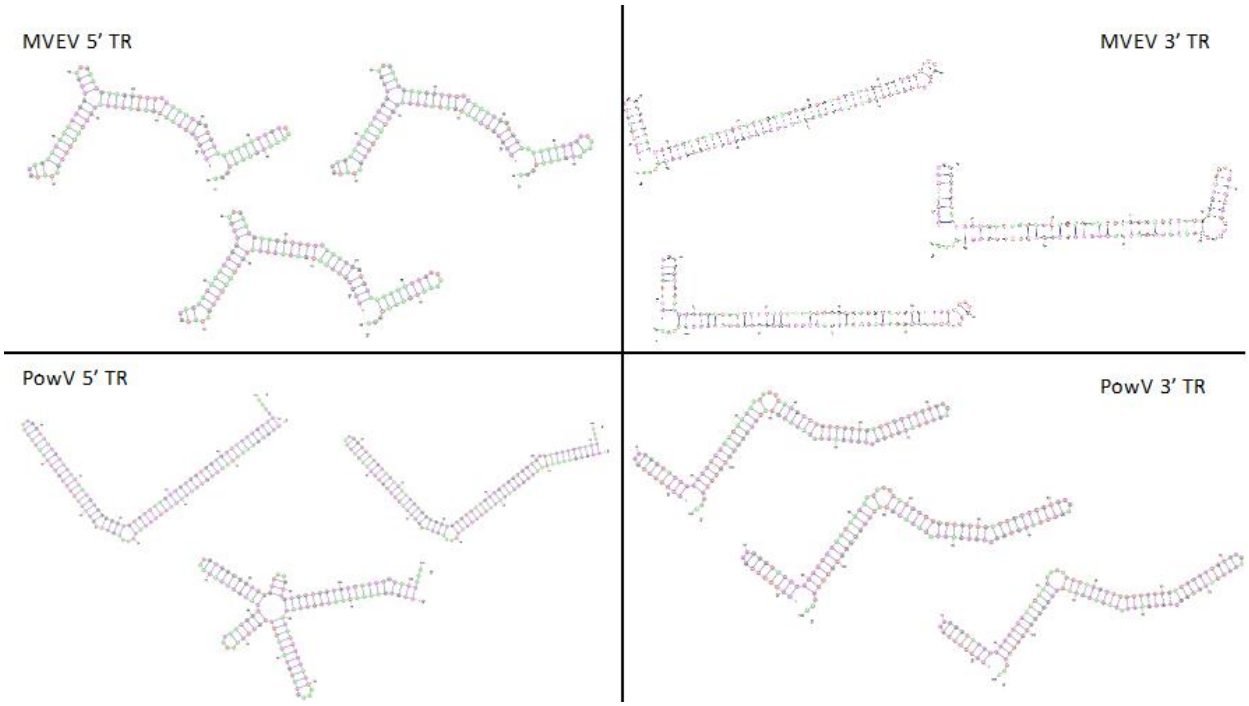


Figure S6.2: The lowest energy structures of MVEV and PowV 5' and 3' terminal regions presented in Figure 6.1. This figure illustrates the next three lowest energy structures for each RNA structure.

6.10 References

1. Fernández-Sanlés, A., Ríos-Marco, P., Romero-López, C. and Berzal-Herranz, A. (2017) Functional information stored in the conserved structural RNA domains of flavivirus genomes. *Frontiers in microbiology*, **8**, 546.
2. Chen, J.-Y., Chen, W.-N., Poon, K.-M.V., Zheng, B.-J., Lin, X., Wang, Y.-X. and Wen, Y.-M. (2009) Interaction between SARS-CoV helicase and a multifunctional cellular protein (Ddx5) revealed by yeast and mammalian cell two-hybrid systems. *Archives of virology*, **154**, 507-512.
3. Zhao, S., Ge, X., Wang, X., Liu, A., Guo, X., Zhou, L., Yu, K. and Yang, H. (2015) The DEAD-box RNA helicase 5 positively regulates the replication of porcine reproductive and respiratory syndrome virus by interacting with viral Nsp9 in vitro. *Virus research*, **195**, 217-224.
4. Li, C., Ge, L.-l., Li, P.-p., Wang, Y., Dai, J.-j., Sun, M.-x., Huang, L., Shen, Z.-q., Hu, X.-c. and Ishag, H. (2014) Cellular DDX3 regulates Japanese encephalitis virus replication by interacting with viral un-translated regions. *Virology*, **449**, 70-81.
5. Selvey, L.A., Dailey, L., Lindsay, M., Armstrong, P., Tobin, S., Koehler, A.P., Markey, P.G. and Smith, D.W. (2014) The changing epidemiology of Murray Valley encephalitis in Australia: the 2011 outbreak and a review of the literature. *PLoS neglected tropical diseases*, **8**, e2656.
6. Mackenzie, J., Lindsay, M., Coelen, R., Broom, A., Hall, R. and Smith, D. (1994) Arboviruses causing human disease in the Australasian zoogeographic region. *Archives of virology*, **136**, 447-467.
7. Floridis, J., McGuinness, S.L., Kurucz, N., Burrow, J.N., Baird, R. and Francis, J.R. (2018) Murray valley encephalitis virus: an ongoing cause of encephalitis in Australia's North. *Tropical medicine and infectious disease*, **3**, 49.
8. Bennett, N. (2008) Murray Valley encephalitis: indeed a ''mysterious disease''. *Vic Inf Dis Bull*, **11**, 94-107.
9. Niven, D.J., Afra, K., Iftinca, M., Tellier, R., Fonseca, K., Kramer, A., Safronetz, D., Holloway, K., Drebot, M. and Johnson, A.S. (2017) Fatal infection with murray valley encephalitis virus imported from Australia to Canada, 2011. *Emerging Infectious Diseases*, **23**, 280.
10. Knox, J., Cowan, R., Doyle, J., Ligtermoet, M., Archer, J., Burrow, J., Tong, S., Currie, B., Mackenzie, J. and Smith, D. (2012) Murray Valley encephalitis: a review of clinical features, diagnosis and treatment. *Medical Journal of Australia*, **196**, 322-326.
11. Kazimírová, M., Thangamani, S., Bartíková, P., Hermance, M., Holíková, V., Štibrániová, I. and Nuttall, P.A. (2017) Tick-borne viruses and biological processes at the tick-host-virus interface. *Frontiers in cellular and infection microbiology*, **7**, 339.
12. Fatmi, S.S., Zehra, R. and Carpenter, D.O. (2017) Powassan virus—a new reemerging tick-borne disease. *Frontiers in public health*, **5**, 342.
13. McLean, D.M., McQueen, E.J., Petite, H.E., MacPherson, L.W., Scholten, T.H. and Ronald, K. (1962) Powassan virus: field investigations in northern Ontario, 1959 to 1961. *Canadian Medical Association Journal*, **86**, 971.
14. Anderson, J.F. and Armstrong, P.M. (2012) Prevalence and genetic characterization of Powassan virus strains infecting Ixodes scapularis in Connecticut. *The American journal of tropical medicine and hygiene*, **87**, 754.
15. Chambers, T.J., Hahn, C.S., Galler, R. and Rice, C.M. (1990) Flavivirus genome organization, expression, and replication. *Annual review of microbiology*, **44**, 649-688.
16. Brinton, M.A., Fernandez, A.V. and Dispoto, J.H. (1986) The 3'-nucleotides of flavivirus genomic RNA form a conserved secondary structure. *Virology*, **153**, 113-121.

17. Brinton, M.A. and Basu, M. (2015) Functions of the 3' and 5' genome RNA regions of members of the genus Flavivirus. *Virus research*, **206**, 108-119.
18. Brinton, M.A. (2014) Replication cycle and molecular biology of the West Nile virus. *Viruses*, **6**, 13-53.
19. Ariumi, Y. (2014) Multiple functions of DDX3 RNA helicase in gene regulation, tumorigenesis, and viral infection. *Frontiers in genetics*, **5**, 423.
20. Li, C., Ge, L.-l., Li, P.-p., Wang, Y., Sun, M.-x., Huang, L., Ishag, H., Di, D.-d., Shen, Z.-q. and Fan, W.-x. (2013) The DEAD-box RNA helicase DDX5 acts as a positive regulator of Japanese encephalitis virus replication by binding to viral 3' UTR. *Antiviral research*, **100**, 487-499.
21. Tingting, P., Caiyun, F., Zhigang, Y., Pengyuan, Y. and Zhenghong, Y. (2006) Subproteomic analysis of the cellular proteins associated with the 3' untranslated region of the hepatitis C virus genome in human liver cells. *Biochemical and biophysical research communications*, **347**, 683-691.
22. Deo, S., Patel, T.R., Chojnowski, G., Koul, A., Dzananovic, E., McEleney, K., Bujnicki, J.M. and McKenna, S.A. (2015) Characterization of the termini of the West Nile virus genome and their interactions with the small isoform of the 2' 5'-oligoadenylate synthetase family. *Journal of structural biology*, **190**, 236-249.
23. Alvarez, D.E., Lodeiro, M.F., Luduena, S.J., Pietrasanta, L.I. and Gamarnik, A.V. (2005) Long-range RNA-RNA interactions circularize the dengue virus genome. *Journal of virology*, **79**, 6631-6643.
24. Villordo, S.M. and Gamarnik, A.V. (2009) Genome cyclization as strategy for flavivirus RNA replication. *Virus research*, **139**, 230-239.
25. Westaway, E. (1987) Flavivirus replication strategy. *Advances in virus research*, **33**, 45-90.
26. Deo, S., Patel, T.R., Dzananovic, E., Booy, E.P., Zeid, K., McEleney, K., Harding, S.E. and McKenna, S.A. (2014) Activation of 2' 5'-oligoadenylate synthetase by stem loops at the 5'-end of the West Nile virus genome. *PloS one*, **9**, e92545.
27. Meier-Stephenson, V., Mrozowich, T., Pham, M. and Patel, T.R. (2018) DEAD-box helicases: the Yin and Yang roles in viral infections. *Biotechnology and Genetic Engineering Reviews*, **34**, 3-32.
28. Valiente-Echeverría, F., Hermoso, M.A. and Soto-Rifo, R. (2015) RNA helicase DDX3: at the crossroad of viral replication and antiviral immunity. *Reviews in medical virology*, **25**, 286-299.
29. Ward, A.M., Bidet, K., Yinglin, A., Ler, S.G., Hogue, K., Blackstock, W., Gunaratne, J. and Garcia-Blanco, M.A. (2011) 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 biology*, **8**, 1173-1186.
30. Dzananovic, E., Chojnowski, G., Deo, S., Booy, E.P., Padilla-Meier, P., McEleney, K., Bujnicki, J.M., Patel, T.R. and McKenna, S.A. (2017) Impact of the structural integrity of the three-way junction of adenovirus VAI RNA on PKR inhibition. *PloS one*, **12**, e0186849.
31. Dzananovic, E., Patel, T.R., Chojnowski, G., Boniecki, M.J., Deo, S., McEleney, K., Harding, S.E., Bujnicki, J.M. and McKenna, S.A. (2014) Solution conformation of adenovirus virus associated RNA-I and its interaction with PKR. *Journal of structural biology*, **185**, 48-57.
32. Dzananovic, E., Patel, T.R., Deo, S., McEleney, K., Stetefeld, J. and McKenna, S.A. (2013) Recognition of viral RNA stem–loops by the tandem double-stranded RNA binding domains of PKR. *Rna*, **19**, 333-344.
33. Chen, Y. and Pollack, L. (2016) SAXS studies of RNA: structures, dynamics, and interactions with partners. *WIREs RNA*, **7**, 512-526.
34. Chen, Y.-L., Lee, T., Elber, R. and Pollack, L. (2019) Conformations of an RNA helix-junction-helix construct revealed by SAXS refinement of MD simulations. *Biophysical journal*, **116**, 19-30.
35. Patel, T.R., Winzor, D.J. and Scott, D.J. (2016) Analytical ultracentrifugation: A versatile tool for the characterisation of macromolecular complexes in solution. *Methods*, **95**, 55-61.

36. Demeler, B. and Gorbet, G.E. (2016) Analytical ultracentrifugation data analysis with UltraScan-III. *Analytical ultracentrifugation: instrumentation, software, and applications*, 119-143.
37. Brosey, C.A. and Tainer, J.A. (2019) Evolving SAXS versatility: solution X-ray scattering for macromolecular architecture, functional landscapes, and integrative structural biology. *Current opinion in structural biology*, **58**, 197-213.
38. Pérez, J. and Vachette, P. (2017) A successful combination: coupling SE-HPLC with SAXS. *Biological small angle scattering: techniques, strategies and tips*, 183-199.
39. Patel, T.R., Chojnowski, G., Koul, A., McKenna, S.A. and Bujnicki, J.M. (2017) Structural studies of RNA-protein complexes: A hybrid approach involving hydrodynamics, scattering, and computational methods. *Methods*, **118**, 146-162.
40. Durand, D., Vivès, C., Cannella, D., Pérez, J., Pebay-Peyroula, E., Vachette, P. and Fieschi, F. (2010) NADPH oxidase activator p67phox behaves in solution as a multidomain protein with semi-flexible linkers. *Journal of structural biology*, **169**, 45-53.
41. Svergun, D. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of applied crystallography*, **25**, 495-503.
42. Svergun, D.I. and Koch, M.H. (2003) Small-angle scattering studies of biological macromolecules in solution. *Reports on Progress in Physics*, **66**, 1735.
43. Svergun, D.I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophysical journal*, **76**, 2879-2886.
44. Volkov, V.V. and Svergun, D.I. (2003) Uniqueness of ab initio shape determination in small-angle scattering. *Journal of applied crystallography*, **36**, 860-864.
45. Parisien, M. and Major, F. (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature*, **452**, 51-55.
46. Darty, K., Denise, A. and Ponty, Y. (2009) VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, **25**, 1974.
47. Svergun, D., Barberato, C. and Koch, M.H. (1995) CRYSOL–a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of applied crystallography*, **28**, 768-773.
48. Uchiyama, S., Noda, M. and Krayukhina, E. (2018) Sedimentation velocity analytical ultracentrifugation for characterization of therapeutic antibodies. *Biophysical Reviews*, **10**, 259-269.
49. Unzai, S. (2018) Analytical ultracentrifugation in structural biology. *Biophysical Reviews*, **10**, 229-233.
50. Kim, D.N., Thiel, B.C., Mrozowich, T., Hennelly, S.P., Hofacker, I.L., Patel, T.R. and Sanbonmatsu, K.Y. (2020) Zinc-finger protein CNBP alters the 3-D structure of lncRNA Braveheart in solution. *Nature communications*, **11**, 148.
51. Krahn, N., Meier, M., To, V., Booy, E.P., McEleney, K., O'Neil, J.D., McKenna, S.A., Patel, T.R. and Stetefeld, J. (2017) Nanoscale assembly of high-mobility group AT-Hook 2 protein with DNA replication fork. *Biophysical journal*, **113**, 2609-2620.
52. Patel, T.R., Meier, M., Li, J., Morris, G., Rowe, A.J. and Stetefeld, J. (2011) T-shaped arrangement of the recombinant agrin G3–IgG Fc protein. *Protein Science*, **20**, 931-940.
53. Reuten, R., Patel, T.R., McDougall, M., Rama, N., Nikodemus, D., Gibert, B., Delcros, J.-G., Prein, C., Meier, M. and Metzger, S. (2016) Structural decoding of netrin-4 reveals a regulatory function towards mature basement membranes. *Nature communications*, **7**, 13515.
54. Kasprzak, W.K. and Shapiro, B.A. (2014) MPGAfold in dengue secondary structure prediction. *Dengue: Methods and Protocols*, 199-224.

55. Liu, X., Liu, Y., Zhang, Q., Zhang, B., Xia, H. and Yuan, Z. (2018) Homologous RNA secondary structure duplications in 3′ untranslated region influence subgenomic RNA production and replication of dengue virus. *Virology*, **524**, 114-126.
56. Ochsenreiter, R., Hofacker, I.L. and Wolfinger, M.T. (2019) Functional RNA structures in the 3′ UTR of tick-borne, insect-specific and no-known-vector flaviviruses. *Viruses*, **11**, 298.
57. Zhang, Y., Zhang, Y., Liu, Z.Y., Cheng, M.L., Ma, J., Wang, Y., Qin, C.F. and Fang, X. (2019) Long non-coding subgenomic flavivirus RNAs have extended 3D structures and are flexible in solution. *EMBO reports*, **20**, e47016.
58. Zettl, T., Mathew, R.S., Shi, X., Doniach, S., Herschlag, D., Harbury, P.A. and Lipfert, J. (2018) Gold nanocrystal labels provide a sequence–to–3D structure map in SAXS reconstructions. *Science Advances*, **4**, eaar4418.
59. Barrows, N.J., Campos, R.K., Liao, K.-C., Prasanth, K.R., Soto-Acosta, R., Yeh, S.-C., Schott-Lerner, G., Pompon, J., Sessions, O.M. and Bradrick, S.S. (2018) Biochemistry and molecular biology of flaviviruses. *Chem Rev*, **118**, 4448-4482.
60. Meyer, A., Freier, M., Schmidt, T., Rostowski, K., Zwoch, J., Lilie, H., Behrens, S.-E. and Friedrich, S. (2020) An RNA thermometer activity of the west Nile virus genomic 3′-terminal stem-loop element modulates viral replication efficiency during host switching. *Viruses*, **12**, 104.
61. Brookes, E., Cao, W. and Demeler, B. (2010) A two-dimensional spectrum analysis for sedimentation velocity experiments of mixtures with heterogeneity in molecular weight and shape. *European Biophysics Journal*, **39**, 405-414.
62. Demeler, B. and Van Holde, K.E. (2004) Sedimentation velocity analysis of highly heterogeneous systems. *Analytical biochemistry*, **335**, 279-288.
63. Brookes, E.H. and Demeler, B. (2007), *Proceedings of the 9th annual conference on Genetic and evolutionary computation*, pp. 361-368.
64. Demeler, B. and Brookes, E. (2008) Monte Carlo analysis of sedimentation experiments. *Colloid and Polymer Science*, **286**, 129-137.
65. Meier, M., Moya-Torres, A., Krahn, N.J., McDougall, M.D., Orriss, G.L., McRae, E.K.S., Booy, E.P., McEleney, K., Patel, T.R. and McKenna, S.A. (2018) Structure and hydrodynamics of a DNA G-quadruplex with a cytosine bulge. *Nucleic acids research*, **46**, 5319-5331.
66. Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H. and Svergun, D.I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of applied crystallography*, **36**, 1277-1282.
67. Rambo, R.P. (2017) ScÅtter, a JAVA-based application for basic analysis of SAXS datasets.
68. Yazdi, M.M., Saran, S., Mrozowich, T., Lehnert, C., Patel, T.R., Sanders, D.A. and Palmer, D.R. (2020) Asparagine-84, a regulatory allosteric site residue, helps maintain the quaternary structure of Campylobacter jejuni dihydrodipicolinate synthase. *Journal of structural biology*, **209**, 107409.
69. Guinier, A., Fournet, G. and Yudowitch, K.L. (1955) Small-angle scattering of X-rays.
70. Patel, T.R., Bernards, C., Meier, M., McEleney, K., Winzor, D.J., Koch, M. and Stetefeld, J. (2014) Structural elucidation of full-length nidogen and the laminin–nidogen complex in solution. *Matrix Biology*, **33**, 60-67.
71. Kozin, M.B. and Svergun, D.I. (2001) Automated matching of high-and low-resolution structural models. *Journal of applied crystallography*, **34**, 33-41.

Chapter 7. Human DDX17 Unwinds Rift Valley Fever Virus Non-Coding RNAs

7.1 Foreword

This chapter consists of a manuscript I wrote in collaboration with Corey R Nelson, whom I share co-authorship with, Sean Park, Simmone D'Souza, Amy Henrickson, Justin Vigar, Dr. Hans-Joachim Wieden, Dr. Raymon Owens, Dr. Borries Demeler and Dr. Trushar R Patel, published in *The International Journal of Molecular Sciences,* Volume 22, issue 1 on December 23, 2020. https://doi.org/10.3390/ijms22010054. This publication focuses on the interaction between DDX17 human RNA helicase and Rift Valley Fever virus non-coding RNA. Corey and I were both involved in all the experiments within the manuscript with him focusing more on the protein aspect, and I the RNA aspect. We both equally contributed to the writing of the final draft. This publication is reproduced with permission from Corey Nelson, MDPI, and the Creative Commons CC by license and re-formatted to fit thesis formatting.

7.2 Author List and Affiliations

Corey R. Nelson[1†], **Tyler Mrozowich**[1†], Sean M. Park[1], Simmone D'souza[2], Amy Henrickson[1], Justin R. J. Vigar[1], Hans-Joachim Wieden[1], Raymond J. Owens[3], Borries Demeler[1,4,5] and Trushar R. Patel[1,2,6]

1. Department of Chemistry and Biochemistry, Alberta RNA Research and Training Institute, University of Lethbridge, 4401 University Drive, Lethbridge, AB T1K 3M4, Canada
2. Department of Microbiology, Immunology and Infectious Disease, Cumming School of Medicine, University of Calgary, Calgary, AB T2N 1N4, Canada
3. Research Complex at Harwell, R92 Rutherford Appleton Laboratories, Harwell, Oxford OX1 0QX, UK
4. Department of Chemistry and Biochemistry, University of Montana, Missoula, MT 59812, USA
5. NorthWest Biophysics Consortium, University of Lethbridge, University of Lethbridge, 4401 University Drive, Lethbridge, AB T1K 3M4, Canada
6. Li Ka Shing Institute of Virology and Discovery Lab, University of Alberta, Edmonton, AB T6G 2E1, Canada

† Equal author contribution.

7.3 Abstract

Rift Valley fever virus (RVFV) is a mosquito-transmitted virus from the Bunyaviridae family that causes high rates of mortality and morbidity in humans and ruminant animals. Previous studies indicated that DEAD-box helicase 17 (DDX17) restricts RVFV replication by recognizing two primary non-coding RNAs in the S-segment of the genome: the intergenic region (IGR) and 5' non-coding region (NCR). However, we lack molecular insights into the direct binding of DDX17 with RVFV non-coding RNAs and information on the unwinding of both non-coding RNAs by DDX17. Therefore, we performed an extensive biophysical analysis of the DDX17 helicase domain ($DDX17_{135–555}$) and RVFV non-coding RNAs, IGR and 5' NCR. The homogeneity studies using analytical ultracentrifugation indicated that $DDX17_{135–555}$, IGR, and 5' NCR are pure. Next, we performed small-angle X-ray scattering experiments, which suggested that DDX17 and both RNAs are homogenous as well. SAXS analysis also demonstrated that DDX17 is globular to an extent, whereas the RNAs adopt an extended conformation in solution. Subsequently, microscale thermophoresis experiments were performed to investigate the direct binding of DDX17 to the non-coding RNAs. The MST experiments demonstrated that DDX17 binds with the IGR and 5' NCR with a dissociation constant of 5.77 ± 0.15 μM and 9.85 ± 0.11 μM, respectively. As $DDX17_{135–555}$ is an RNA helicase, we next determined if it could unwind IGR and NCR. We developed a helicase assay using MST and fluorescently-labeled oligos, which suggested $DDX17_{135–555}$ can unwind both RNAs. Overall, our study provides direct evidence of $DDX17_{135–555}$ interacting with and unwinding RVFV non-coding regions.

7.4 Introduction

Rift Valley fever virus is part of the Bunyaviridae family and the genus Phlebovirus. The virus was first identified in the early 1930s during a large outbreak on a sheep farm in the Rift Valley of Kenya (1). Since then, the virus transmission has been reported in several countries located within Sub-Saharan Africa and the Arabian Peninsula due to infected livestock trade. Transmission of this virus is through competent mosquito vector hosts, *Aedes*

and *Culex*, to animals and humans. Infection is currently untreatable and primarily affects domesticated animals such as camels, goats, sheep, and cattle, but can also affect humans (2). Infections in animals mainly occur through mosquito bites, and terminal human hosts can be infected by infected mosquitos and direct contact with infected ruminant blood and bodily fluids. Symptoms of RVFV infection in humans can include acute febrile illness followed by hemorrhagic fever, encephalitis, or ocular disease (2). These mosquito vectors can also transmit a variety of flaviviral diseases such as Zika, yellow fever, and chikungunya. Additionally, its presence on every continent except for Antarctica makes this virus a severe threat to global health and food security (3). There is currently an inactivated vaccine (MP-12) that has been shown to confer long-term immunity in humans with a single dose. This vaccine is also effective in animals; however, it is not currently commercially available (4).

According to the World Health Organization (WHO), RVFV infection is one of the top eight emerging diseases likely to cause significant epidemics that currently have no medical countermeasures (5). RVFV is maintained in the environment via vertical transmission from the mosquito vector to offspring during periods of high rainfall, which amplifies mosquito breeding (6). Infectious outbreaks occur after long intervals of dormancy, between 5–15 years, but can cause detrimental economic losses due to live-stock infection, which causes mortality rates of 10–30% and >90% abortion rates. Moreover, these epizootic outbreaks have resulted in a total death toll of over 100,000 sheep, over half a million livestock abortions, and more than 2300 human deaths (1,7).

RVFV is an enveloped virus that contains a linear, tripartite, ssRNA ambisense genome (8). The total tripartite genome size for RVFV is 10.4 kb and encompasses three different viral RNA components which are replicated in the host cell cytoplasm: L (large, 6.4 kb) and M (medium, 2.3 kb), which are both negative sense, and S (small, 1.7 kb) which is ambisense (9). The L-segment encodes the RNA-dependent RNA polymerase, while the M-segment encodes two envelope glycoproteins, Gn and Gc, and two accessory proteins. The S-segment

(S2) ambisense RNA encodes a positive-sense RNA template of the non-structural protein (NS) and a negative-sense RNA for the viral nucleoprotein (N) (10,11). The S-segment contains two notable non-coding regions; the 5′ non-coding region, responsible for transcription and translation initiation of the NS segment, and the intergenic region, responsible for transcription termination on both the NSs and N mRNA (12). These two regions have been demonstrated to form hairpin-like structures, which are recognized by host cellular machinery. The human protein, DDX17, which is a DEAD-box helicase, has been shown to interact with the 5′ non-coding S-segment (RVFV NCR) and the non-coding sequence between N and NSs (RVFV IGR) (11). The knockdown of DDX17, but not its paralog DDX5, led to unrestricted viral replication of RVFV in U2OS cells (11).

DEAD-box helicases have been described as ATP-dependent chaperones that reconfigure RNA by disrupting secondary and tertiary RNA–RNA or RNA–protein interactions (13,14). The DDX17 helicase has roles in transcription, splicing, mRNA decay, rRNA biogenesis, and miRNA processing as well as antiviral defense (15,16). The target motifs on the viral RNA S-segment are two hairpin structures, which are unique to the Bunyaviridae family (10). CLIP-seq data indicate that DDX17 interacts with both the IGR and 5′ NCR of the RVFV S-segment. We wanted to examine these interactions in vitro to determine if DDX17 is capable of functioning independently or whether it requires additional binding partners, as described previously (17). Additionally, we would like to investigate the structural differences between the non-coding RNAs (ncRNAs) and how this could affect their interactions with DDX17.

Therefore, by utilizing multiple biophysical techniques, we have characterized the interaction between RVFV non-coding RNA (ncRNA) and $DDX17_{135–555}$ to determine that not only does $DDX17_{135–555}$ directly interact with RVFV ncRNA, it also unwinds the ncRNA in the presence of ATP. This work supports previous observations of direct antiviral effects of DDX17 (11) while providing a new, easy approach to investigate the helicase activity of a protein.

7.5 Results

7.5.1 Purification of DDX17$_{135–555}$, RVFV S-segment IGR, and 5′NCR

DDX17$_{135–555}$ was overexpressed in *Escherichia coli* (Lemo21) and purified using affinity and size exclusion chromatography (SEC), as detailed in the Materials and Methods section. Figure 7.1A shows the schematics of full-length DDX17 and the truncated DDX17$_{135–555}$ that was used in this study, as we were unable to express sufficient amounts of full-length DDX17. As presented in Figure 7.1B, the peak fractions (15 mL to 17 mL), devoid of any contamination or aggregation, were collected, followed by the purity check using SDS-PAGE. As presented in the inset to Figure 7.1B, the final preparation does not contain any degraded material, and it corresponds to the correct molecular weight of ~50 kDa. To further study the homogeneity

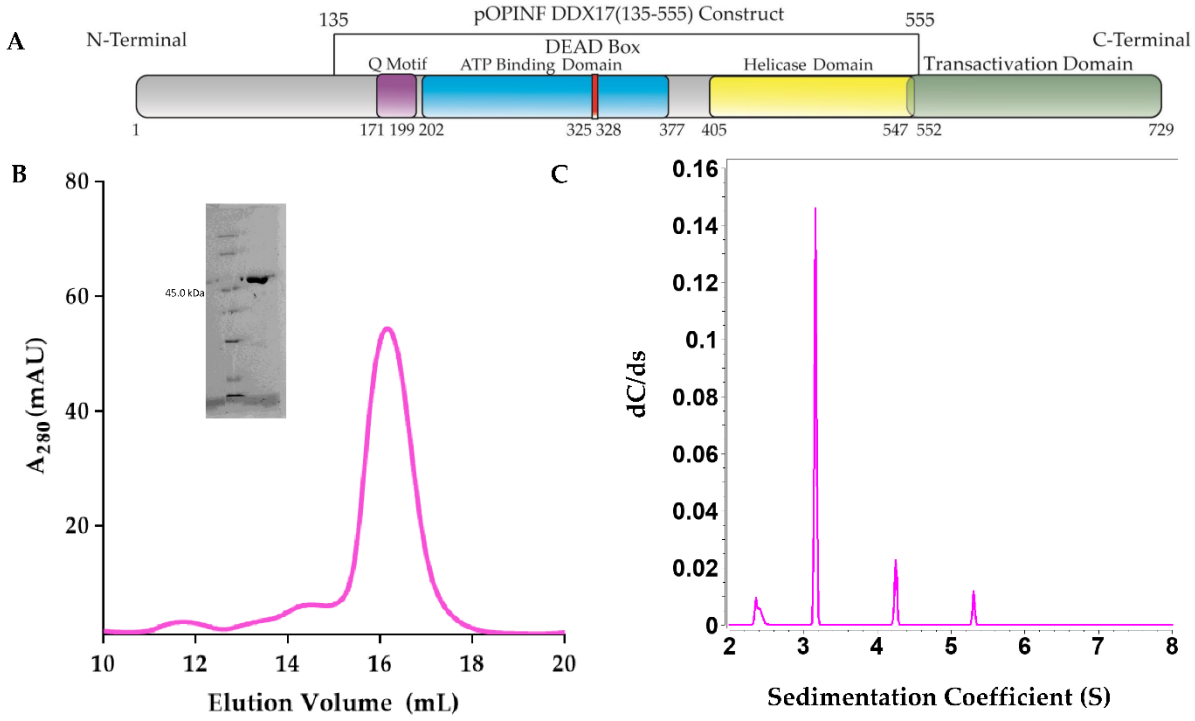


Figure 7.1. Purification of DDX17$_{135–555}$ (**A**) Schematic representation of DDX17 highlighting individual domains. DDX17$_{135–555}$, which contains the Q motif, ATP binding domain, DEAD-box, and the helicase domain, was used in downstream studies. (**B**) The chromatogram from the size exclusion purification (Superdex 200 Increase gl 10/300) of DDX17$_{135–555}$, suggesting that DDX17$_{(135–555)}$ can be purified to ~68% homogeneity (~16 mL). The y-axis represents absorbance at 260 nm while the x-axis represents elution volume. We collected peak fractions from 15.5 to 16.5 mL for subsequent analysis. The inset to Figure 7.1B represents the SDS-PAGE analysis of DDX17$_{135–555}$ (48.45 kDa) following size exclusion chromatography. **C)** Sedimentation coefficient distribution of DDX17$_{135–555}$ obtained from analytical ultracentrifugation sedimentation velocity (SV-AUC) experiment. The peak at ~3.16S represents monodispersed DDX17$_{135–555}$. Sedimentation coefficient values are corrected to standard solvent conditions (20 °C in water).

of DDX17$_{135–555}$ in solution, the SEC-purified preparation that presented a single band in SDS-PAGE was used to perform an analytical ultracentrifugation sedimentation velocity experiment (SV-AUC). The SV-AUC results suggest that DDX17$_{135–555}$ is mainly homogenous with a sedimentation coefficient of 3.16 S (Figure 7.1C) and a diffusion coefficient of 5.22 × 10^{-7} cm^2/s (Table 7.1).

Table 7.1. Solution properties of DDX17$_{135–555}$, IGR, and 5′ NCR.

Sample	DDX17$_{135_555}$	IGR	5′ NCR
M_w (kDa, sequence)	48.45	23.82	24.70
Sedimentation coefficient, S (10^{-13} s) ▽	3.16	4.07	4.18
Diffusion coefficient D (10^{-7} cm^2/s) ▽	5.22	7.62	6.58
R_h (Å) ▽	41.06	28.11	32.64
I(0) #	0.003 ± 2.5 × 10^{-5}	0.026 ± 4.4 × 10^{-5}	0.022 ± 2.6 × 10^{-4}
q.Rg range #	0.39–1.30	0.26–1.29	0.40–1.29
R_g (Å) #	24.78 ± 0.36	36.42 ± 0.10	50.44 ± 0.88
I(0) Δ	0.003 ± 2.3 × 10^{-5}	0.026 ± 4.3 × 10^{-5}	0.019 ± 1.7 × 10^{-4}
R_g (Å) Δ	25.46 ± 0.27	38.00 ± 0.08	46.66 ± 0.34
D_{max} (Å)Δ	79.21	122	148
X^2 *	~1.00	~1.10	~1.30
NSD *	0.52 ± 0.02	0.73 ± 0.02	0.58 ± 0.01

The M_w values were calculated using nucleotide sequences. ▽—determined using SV-AUC analysis and UltraScan-III package (18). Sedimentation coefficients obtained following genetic algorithm–Monte Carlo analysis. #—obtained from Guinier analysis (21). Δ—determined using $P(r)$ analysis using the GNOM program (24). *—values derived from DAMMIN (26) and DAMAVER (27) analysis.

The in vitro transcribed RVFV IGR and 5′ NCR RNAs were purified using SEC, similar to DDX17$_{135–555}$ (Figure 7.2A). The IGR eluted at approximately ~14 mL, while the 5′ NCR eluted at ~14.5 mL. Peak fractions were collected and analyzed by urea-PAGE, which displayed a single band (Figure 7.2A inset). Next, we utilized SV-AUC to determine the purity of SEC-purified RVFV ncRNA. Our SV-AUC analysis suggested that monomeric IGR and 5′ NCR have sedimentation coefficients of 4.07 S and 4.18 S, respectively. The SV-AUC analysis also yielded diffusion coefficients of 7.62 × 10^{-7} cm^2/s and 6.58 × 10^{-7} cm^2/s, respectively. Overall, both ncRNAs appear to be relatively pure (Figure 7.2B).

7.5.2 Solution Conformation of DDX17$_{135–555}$, RVFV S-Segment IGR, and 5′NCR

SAXS analysis allows for low-resolution structural determination of biomolecules in solution. The instrumentation provided at the B21 Beamline (Diamond Light Source, UK) allows for the employment of HPLC connected in-line to SAXS detection to maintain confidence in the monodispersity of samples, keeping them free of aggregates and degradation (19-21). SEC-SAXS data for the merged datasets are presented in Figure 7.3A. The merged data were further processed using Guinier analysis (plot of (I(q)) vs. (q2)) to detect the purity and for the determination of the R_g (average root mean squared radius from the center of mass for the biomolecule) from the low-q region (22). Figure 7.3B represents the Guinier plots for IGR, 5' NCR, and $DDX17_{135-555}$, whereas the linearity of the low-q region indicates that all three biomolecules were monodisperse. R_g values of 36.42 ± 0.10, 50.44 ± 0.88, and 24.78 ± 0.36 for IGR, 5' NCR, and $DDX17_{135-555}$, respectively, were obtained from Guinier analysis (see

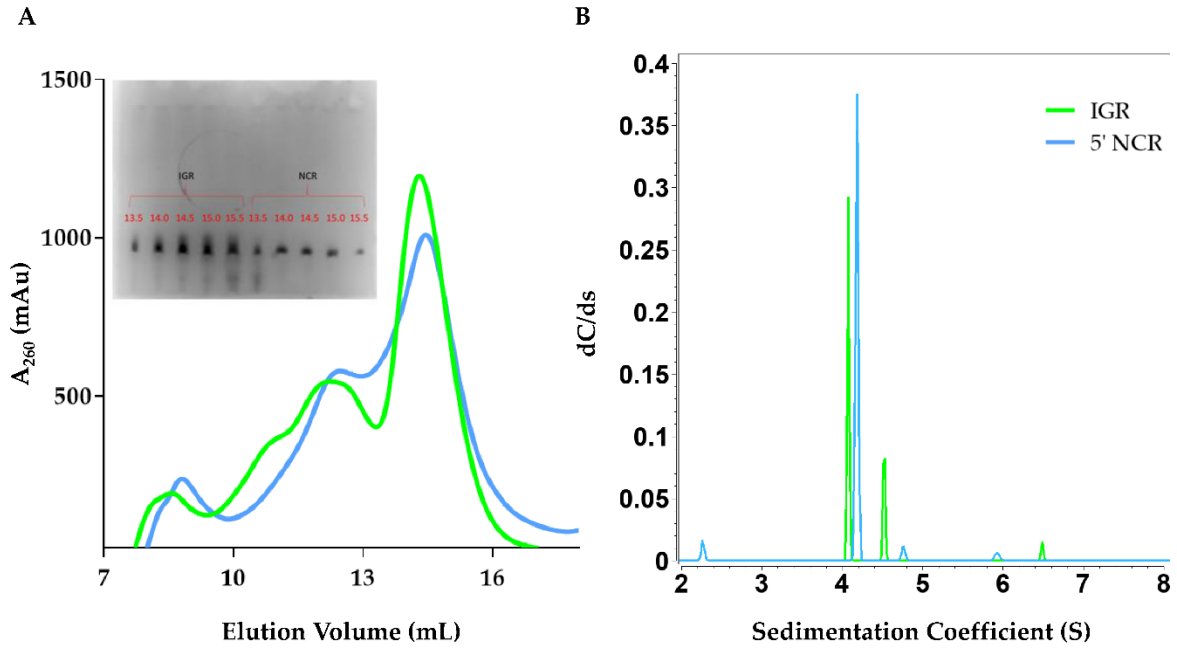


Figure 7.2. Purification and hydrodynamic characterization of in vitro transcribed Rift Valley fever virus RNA. (A) Size exclusion chromatogram of the elution profile of both Rift Valley fever virus (RVFV) 5' intergenic region (IGR) and RVFV 5' non-coding region (NCR). The y-axis represents absorbance at 260 nm while the x-axis represents elution volume. An inset to Figure 7.2A represents the urea-PAGE (7.5%) analysis of RVFV IGR and 5' NCR after size exclusion chromatography. Each well represents 10 µL of a 500 µL elution fraction from size exclusion chromatography. The gel was run for 25 min, at 300 V in 0.5× TBE (Tris-Borate-EDTA) running buffer and was visualized using Sybr Safe dye. (B) Sedi-mentation coefficient distribution profiles for RVFV 5' IGR (green) and RVFV 5' NCR (blue) from SV-AUC. The primary SV peaks for each RNA are 4.07 S and 4.18 S for IGR and 5' NCR, respectively, and represent the monomeric form. Sedimentation coefficient values were corrected to standard conditions (20 °C in water).

Table 7.1). After we confirmed monodispersity from Guinier analysis, we further processed the SAXS scattering data from Figure 7.3A to obtain dimensionless Kratky plots (20-23) which allowed for analysis of the foldedness of the biomolecules (Figure 7.3C). In general, globular biomolecules in solution show a well-defined maximum value of 1.1 at $q*R_g = 1.73$ (24). The dimensionless Kratky plots for the two ncRNAs suggested that both are well folded and extended in solution, whereas $DDX17_{135–555}$ is relatively more compact.

Next, indirect Fourier transformations on each dataset were performed to convert the reciprocal-space information of data presented in Figure 7.3A to real-space electron pair distance distribution functions (P(r)) plots, which are presented in Figure 7.3D using the

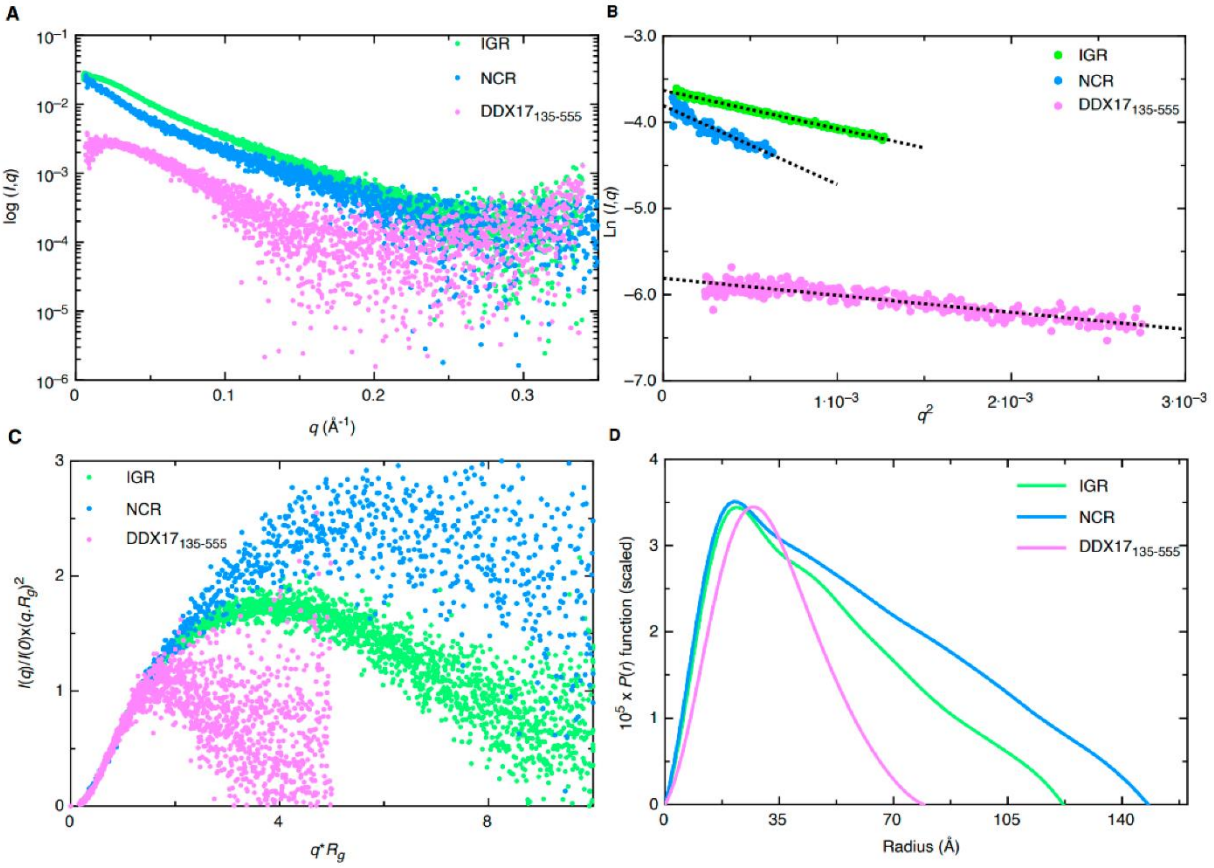


Figure 7.3. Small-angle X-ray scattering (SAXS) characterization of RVFV RNA (IGR and NCR) and $DDX17_{135–555}$. (A) Merged scattering data of RVFV RNA and $DDX17_{135–555}$ showing scattering intensity (log I(q)) vs. scattering angle ($q = 4\pi\sin\theta/\lambda$). **(B)** Guinier plots allowing for the determination of R_g from low-angle region data and representing the homogeneity of samples. **(C)** Dimensionless Kratky plots $(I(q)/I(0)*(q*R_g)^2$ vs. $q*R_g)$ of RVFV RNA and $DDX17_{135–555}$, demonstrating extended structures for RVFV RNA and a more compact structure for $DDX17_{135–555}$. **(D)** Pair distance distribution (P(r)) plots for RVFV RNA and $DDX17_{135–555}$ which allow for the determination of R_g from the entire SAXS dataset, and maximal particle dimension (D_{max}).

GNOM (25) program. Using the P(r) plots, the R_g was obtained along with the D_{max} (maximal particle dimension) for all three biomolecules. Importantly, compared to Guinier analysis, which provides R_g from the low-q region, the P(r) analysis utilizes a larger range of the dataset which adds to the reliable determination of the R_g and D_{max}. Table 7.1 contains all values calculated from the P(r) analysis; we obtained a D_{max} of ~120 Å, 145 Å, and 80 Å for IGR, 5′ NCR, and $DDX17_{135–555}$, respectively. Additionally, we obtained P(r) R_g values of 38.00 ± 0.08, 46.66 ± 0.34, and 25.46 ± 0.27 Å for IGR, 5′ NCR, and $DDX17_{135–555}$, respectively. These values correlate very well to those obtained from prior Guinier analysis, indicating these data are suitable to proceed with low-resolution structure determination. The P(r) plot is also indicative of a biomolecules' relative solution conformation; a more globular-shaped biomolecule will adopt a bell-shaped P(r) distribution with a maximum at $D_{max}/2$ (26), and a more extended molecule will adopt a bell-shaped curve with an extended tail, suggesting an

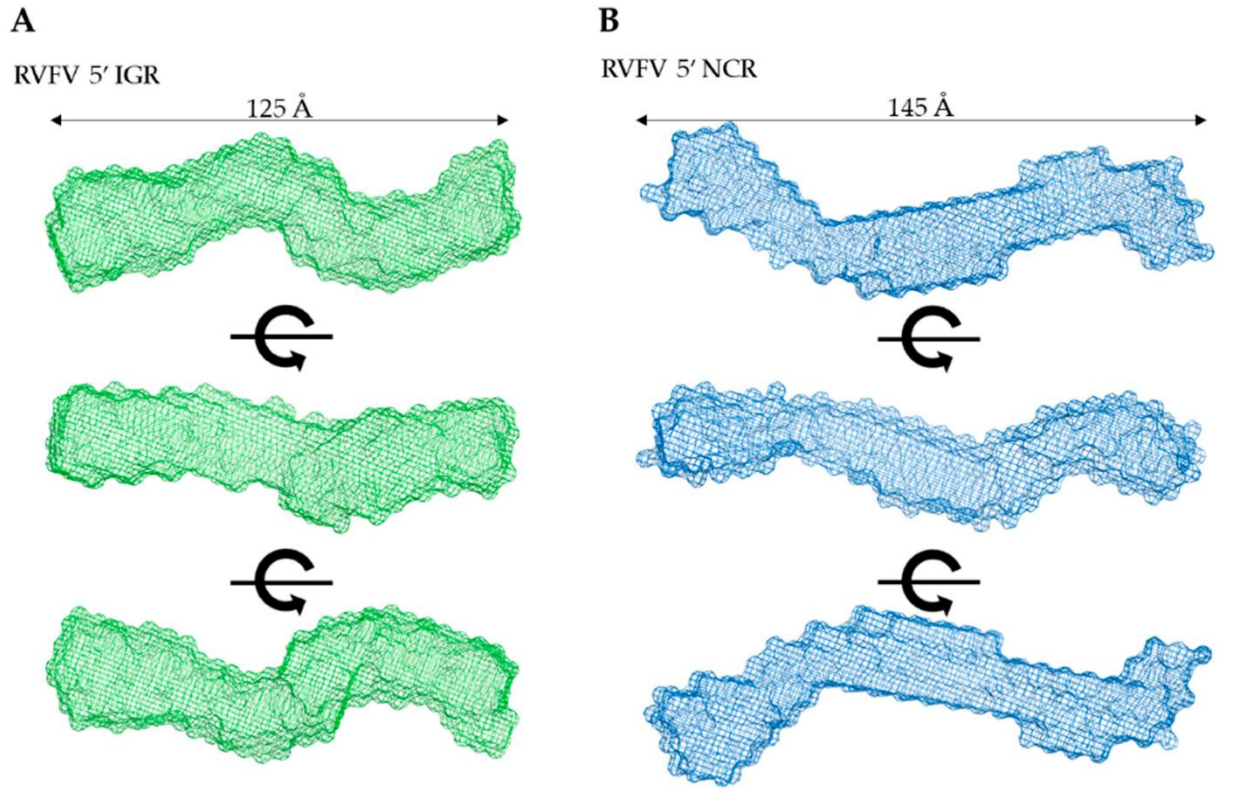


Figure 7.4. Low-resolution structure determination via SAXS for RVFV 5' NCR and RVFV 5' IGR, indicating that these RNA molecules adopt an extended solution structure. (A,B) Three structures representing sequential 90° rotational angles from the top panel structure. Dimensions represent the D_{max} obtained from P(r) analysis.

elongated structure (21). The P(r) plot for $DDX17_{135–555}$ adopts a typical bell-shaped curve, which suggests that this protein is more globular relative to the ncRNAs (Figure 7.3D).

Next, we employed DAMMIN (27) to obtain low-resolution structures for each biomolecule, which involves a simulated annealing protocol allowing for the incorporation of P(r) data (D_{max} and R_g as constraints). Twelve models were calculated for all three biomolecules and all models have excellent agreement (X^2) between the experimentally obtained scattering data and the calculated scattering data (Table 7.1). Following DAMMIN, we employed DAMAVER (28) for the alignment and rotation of all 12 models to gain an averaged filtered structure for each biomolecule, which represents aver-aged structural features from individual models (Figures 4 and 5A) (28). For each case, the overlap function, the normalized spatial discrepancy, was estimated to provide a measure of the goodness of fit of the superimposition of each model. Table 7.1 presents the NSD values for the 12 models calculated for each biomolecule, and the low values suggest that the models in each case are

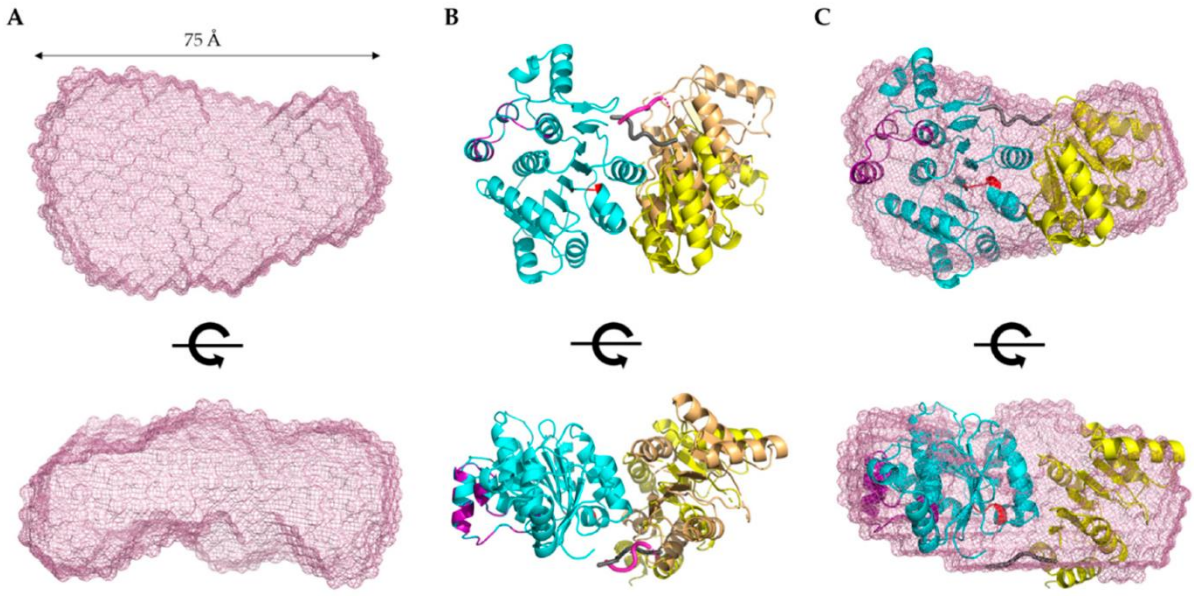


Figure 7.5. Structural modeling of $DDX17_{135–555}$. (A) Low-resolution SAXS structure indicating that $DDX17_{135–555}$ adopts an extended globular conformation in solution. The bottom panel represents a 90° rotation of the x-axis from the top panel. Dimensions represent the D_{max} from P(r) analysis. **(B)** CORAL-derived models of $DDX17_{135–555}$, suggesting a linker (purple/gray chain) between the ATP-binding domain (blue ribbon), and the helicase domain (yellow/brown ribbon), allowing them to adopt different orientations. **(C)** SAXS envelope overlayed with the CORAL-derived representative model highlighting an agreement between high- and low-resolution models.

highly similar to each other. The models presented in Figures 4 and 5A are the averaged filtered structures for NCR, 5' IGR, and DDX17$_{135–555}$, which indicate that both ncRNAs adopt extended structures in solution, while DDX17$_{135–555}$ has a nearly globular conformation.

Recently, a high-resolution structure of DDX17 containing the ATP-binding and helicase domains (6UV0) was determined using X-ray crystallography (29). We noticed a flexible linker between the ATP-binding and helicase domains, which could not be re-solved in the high-resolution crystal structure. Therefore, we sought to use the scattering data of DDX17$_{135–555}$ to perform high-resolution modeling using the program CORAL, as described elsewhere (30). Using the crystal structure's high-resolution in-formation of the ATP-binding domain (155aa-382) and the helicase domain (389aa-555), we calculated 12 separate models and assessed their quality by comparing model-derived SAXS data with experimentally collected SAXS data. Each of the 12 models we calculated has χ^2 values of ~1.2, suggesting they are a good fit for the original data. This led us to believe that the helicase domain can

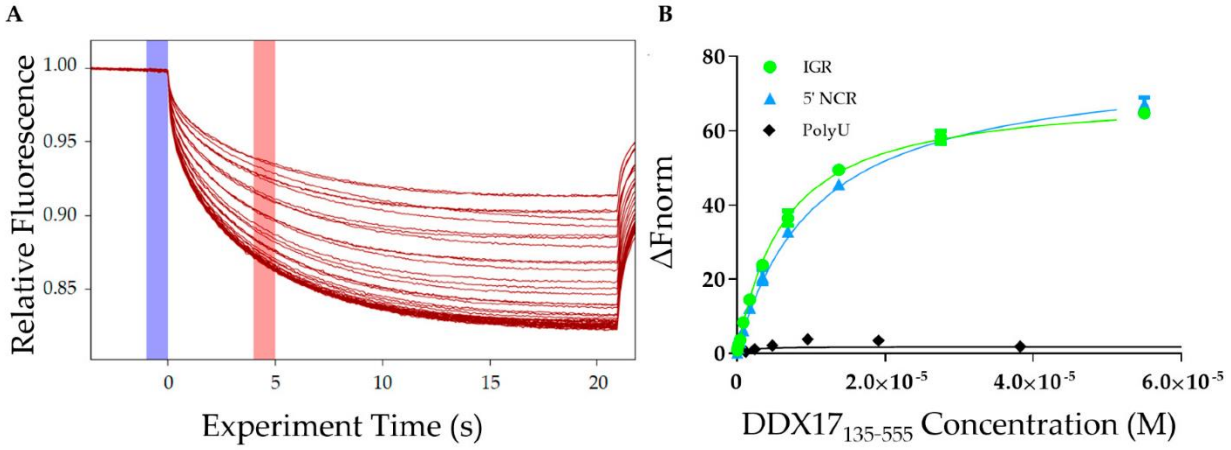


Figure 7.6. Interaction studies of DDX17135–555 with IGR and 5' NCR. (A) Microscale thermophoresis (MST) traces indicating the change in fluorescence when exposed to the infrared laser. Each trace represents a different concentration of DDX17$_{135–555}$ and is used to assess how the change in concentration affects the fluorescently labeled RNA migration. The blue highlight is the "cold" region and the red highlight is the "hot" region. The difference between these regions is used to calculate the ΔF_{norm}. **(B)** The MST binding curves for the IGR and 5' NCR RVFV RNAs (n = 3). RNA was used at a concentration of 40 nM while DDX17$_{135–555}$ was titrated up to a maximum concentration of 55µM. The y-axis ΔF_{norm} is the change in fluorescent migration normalized to 0. The dissociation constant for DDX17$_{135–555}$ and the IGR was determined to be 5.78 ± 0.15 µM (reduced X2 = 0.967, Std. error of regression = 0.702) while for DDX17$_{135–555}$ and the 5' NCR was determined to be 9.85 ± 0.11 µM (reduced X2 = 0.996, Std. error of regression = 0.351). We used polyU RNA as a negative control (black diamonds) that did not bind to DDX17$_{135–555}$.

adopt multiple different orientations in solutions, consistent with our initial low-resolution SAXS structure presented in Figure 7.5A. Figure 7.5B presents the CORAL-derived representative models, which high-light the relative orientations of the helicase domain due to the presence of a linker. Figure 7.5C demonstrates the overlay of the CORAL-derived model with the DDX17$_{135–555}$ low-resolution structure, indicating an overall agreement between both approaches.

7.5.3 DDX17 Binds to the IGR and 5'NCR Non-Coding RNAs

After analyzing the homogeneity of DDX17$_{135–555}$, RVFV S-segment IGR, and 5' NCR, we determined the affinity of DDX17$_{135–555}$ for both ncRNAs using microscale thermophoresis (MST). MST is a powerful technique that allows for rapid interaction analysis by measuring the change in fluorescent migration as the molecules are excited via infrared laser (31,32). DDX17$_{135–555}$ was titrated against the fluorescently labeled RVFV RNAs. The addition of

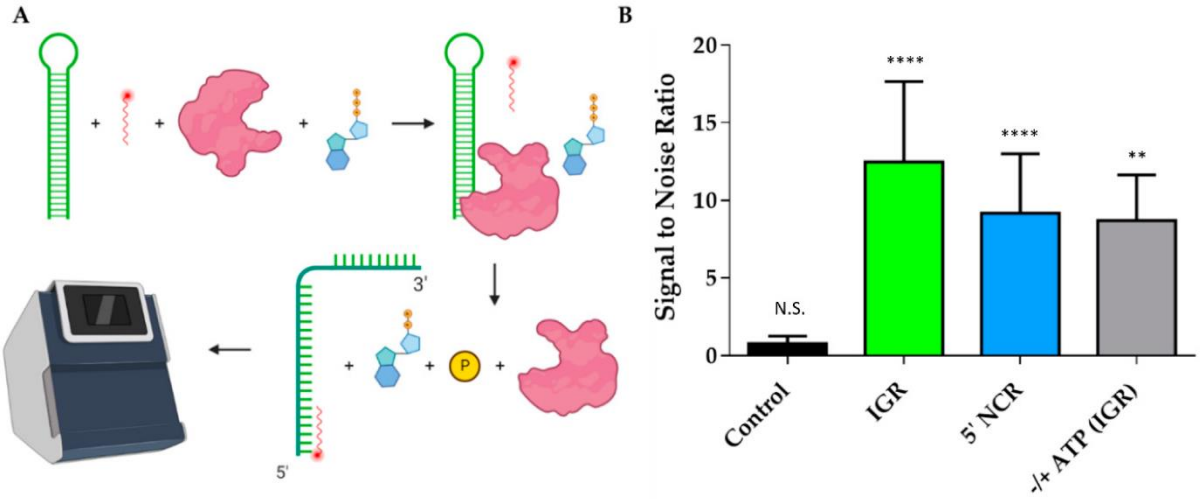


Figure 7.7. DDX17135–555 helicase assays performed using MST. (A) Representation of helicase assay using MST. The helicase assay was performed by combining the RNAs, fluorescently labeled DNA oligo, DDX17$_{135–555}$, and ATP. DDX17$_{135–555}$ hydrolyzes the ATP and unwinds the RNA, giving the oligo access to the newly opened complimentary site. The fluorescently labeled DNA oligo hybridized to the RNA can be measured in the MST by detecting the change in migration. This is compared to a control that uses bovine serum albumin (BSA) instead of DDX17$_{135–555}$, and a change in the migration indicates that the RNA has been unwound by DDX17. **(B)** Signal to noise ratios of different comparative assays (n = 3). Control compared a reaction mix without protein to BSA, signal to noise did not meet the threshold of 5 and was not significant, unpaired t-test ($p = 0.9350$, N.S.). DDX17$_{135–555}$ caused a significant change in the IGR ($p < 0.0001$, ****), with the signal to noise ratio reaching 12.5. The 5' NCR also experienced a significant shift in the presence of DDX17135–555 ($p < 0.0001$, ****), having a signal to noise ratio of 9.17. Gray bar represents including ATP vs. not including ATP in the reaction mixture, showing ATP causes a significant change in fluorescent migration, having a signal to noise of 8.70 ($p = 0.0059$, **).

$DDX17_{135–555}$ (the ligand) to the fluorescent RNA molecules (the target) causes them to migrate at a rate different than when $DDX17_{135–555}$ is absent. A dissociation constant is determined by relating the change in fluorescent migration of the target to the concentration of the added ligand (33). Figure 7.6A represents MST traces, where the blue highlight represents the "cold" region which is used to normalize the change of fluorescence measured in red, representing the "hot" region. Our MST studies demonstrate that $DDX17_{135–555}$ interacts with IGR and 5′ NCR with dissociation constants of 5.77 ± 0.15 μM and 9.85 ± 0.11 μM, respectively (Figure 7.6B).

7.5.4 DDX17 Unwinds RVFV RNA in an ATP-Dependent Fashion

Since we confirmed that $DDX17_{135–555}$ binds to both ncRNAs, we wanted to evaluate DDX17's ability to unwind the RNAs. Figure 7.7A is a schematic representation of the experimental design which describes the overall approach of utilizing MST to perform a helicase assay. The signal to noise ratio, which is a measure of significance that uses the response amplitude of the MST traces, is indicated in Figure 7.7B. To assess statistical significance, we used unpaired t-tests. We determined that the fluorescent migration did not experience a significant change ($p = 0.9350$, signal to noise = 0.750) in the presence of BSA with the reaction mixture (RNA + fluorescent oligo + ATP), suggesting that BSA cannot unwind RNA, which makes it a suitable control for the subsequent experiments. Next, we compared the BSA reaction mixture to the reaction mixture with $DDX17_{135–555}$. The results suggest that the addition of DDX17 with either IGR or 5′ NCR causes a significant change in the migration of fluorescence ($p < 0.0001$ for both, signal to noise = 12.5 and 9.17, respectively), indicating a binding event occurred upon the addition of DDX17. Collectively, our analysis demonstrates that $DDX17_{135–555}$ can unwind the RNA, allowing the hybridization of the DNA oligo to the RNA(s). To determine the effect of ATP on the helicase activity of DDX17, we compared the reaction mix with and without ATP (gray bar). We observed that the presence of ATP resulted

in a significant difference in fluorescence migration compared to without ATP ($p = 0.0059$, signal to noise = 8.70).

7.6 Discussion

The study performed by Moy et al. in 2014 (11) concluded, *in vivo*, that U2OS human cells infected with RVFV cause activation of DDX17 to restrict RVFV replication through an interferon-independent pathway. CLIP-seq analysis determined that DDX17 binds to two essential stem-loop regions on the RVFV S-segment RNA: IGR and 5′ NCR (11). We, therefore, sought to characterize this interaction in vitro to substantiate that DDX17 is an interacting partner of RVFV ncRNAs.

We expressed and purified a construct that contains both the ATP-binding domain and the helicase domain, $DDX17_{135–555}$ (Figure 7.1). Next, we transcribed, purified, and characterized the RVFV IGR and 5′ NCR ncRNAs in vitro. As AUC is a reliable and widely accepted technique to assess the solution state of biomolecules (34-36), we performed the SV-AUC experiments. The SV-AUC data suggested that both ncRNAs are relatively pure, with the presence of dimer and tetrameric assemblies (Figure 7.2B), which is similar to our prior study on Murrey Valley and Powassan virus ncRNAs where we also observed the presence of oligomeric species (21). Similarly, SV-AUC studies also indicated that $DDX17_{135–555}$ is mainly monomeric at the examined concentration. We also obtained the diffusion coefficients and the Stokes radii for $DDX17_{135–555}$, IGR, and 5′ NCR (Table 7.1).

SAXS excels at being a complementary structural biophysical method by enabling solution structure studies of virtually all biomolecules, and their biomolecular complexes (19,23,30,37-40). While SAXS structures are low resolution in comparison to high-resolution structures determined using X-ray crystallography or NMR, oftentimes obtaining high-quality crystals for crystallography or biomolecular labeling for NMR is challenging (19,38,40-43). By employing HPLC-SAXS for data collection instead of traditional SAXS, we ensure that our

collected scattering data will be monodispersed. These monodispersed preparations were confirmed by the linearity of fit in the low-q region using the Guinier analysis (Figure 7.3B). Using Guinier analysis, we also calculated R_g values for all three biomolecules (based on low-q region) (Figure 7.3B) and compared them to those calculated through P(r) analysis (Figure 7.3D). The R_g values for both analyses were highly similar (Table 7.1), which confirms that our data are reliable and it is worth proceeding with more analysis. Dimensionless Kratky analysis suggested that the IGR and 5′ NCR adopt an elongated structure (Figure 7.3C). Finally, the P(r) distribution (Figure 7.3D) reveals that both ncRNAs quickly increase to the maxima, and then steadily decrease, which suggests an elongated structure, as observed earlier (21,38). Comparatively, the P(r) distribution of $DDX17_{135–555}$ displays a skewed Gaussian distribution, suggesting that it adopts a more compact conformation compared to the ncRNAs (Figure 7.3D). We observe that both ncRNAs have different D_{max} (110 vs. 145 Å), despite having a similar length. The 5′ NCR, based on its D_{max}, likely contains extended amounts of single-stranded regions (Figure 7.4A), whereas IGR could have a higher content of double-stranded structures (Figure 7.4B). IGR and 5′ NCR (Figure 7.4A,B) con-firm that both RNAs adopt an elongated structure, as indicated by initial dimensionless Kratky analysis. The ratio of R_g to R_h is a good indicator of the solution conformation of biomolecules. Compact spherical biomolecules typically have an R_g/R_h ratio of ~0.70. This ratio increases as the shape of the molecule changes from globular to extended conformation (44,45). For IGR and 5′ NCR, the Rg/Rh values are 1.35 and 1.43, suggesting that both ncRNAs have extended conformations. For $DDX17_{135–555}$, we obtained an R_g/R_h of 0.62, indicating that it is more globular than the ncRNAs.

The low-resolution structural modeling of DDX17 confirmed its extended globular nature (Figure 7.4A). A secondary strength of SAXS is the ability to combine high-resolution structures or homology models of individual domains, or computational studies with low-resolution SAXS models (19,30,38). The crystal structure of DDX17 containing the ATP-

binding and helicase domain (6UV0) was determined (29), allowing us to compare their high-resolution data to our low-resolution models to evaluate the validity of our models. Since the flexible linker between the ATP-binding and helicase domains was not resolved, we performed structural modeling using CORAL, which suggested that relative to the ATPase domain, the helicase domain exhibits conformational flexibility in solution (Figure 7.5C).

To establish the direct interaction between $DDX17_{135–555}$ and both IGR and 5′ NCR ncRNAs, we performed MST assays as described previously (30,46,47). Our analysis indicated that both RNAs interact with $DDX17_{135–555}$. However, despite having relatively similar nucleotide length, IGR binds with a comparatively higher affinity to 5′ NCR (5.77 μM for the IGR vs. 9.85 μM for the 5′ NCR) (Figure 7.6B). Compared to the observations made for DDX5 (a DDX17 homolog), our results suggest that the DDX17 interacts with RVFV RNAs weakly (in μM range) (48,49). However, an important distinction between previous studies and our work is that we have used considerably longer ncRNAs, and the minimalistic DDX17 construct. This could result in differences in specificity, nonetheless, we have demonstrated that our construct is specific to the RVFV RNAs and binds with them with different affinities. Considering how compact the IGR is, based on scattering analysis, it may indicate that DDX17 has tighter binding to double-stranded RNA regions. Although DDX17 is primarily located in the nucleus, its presence in the cytoplasm and ability to interact with RNAs, including the RVFV ncRNAs, suggests that DDX17 may act as a sensor for these viral RNAs within the cytoplasm (50), similar to other helicases and host proteins, like DDX3X and Protein Kinase R (PKR) (51,52). Since DDX17 is a known helicase, we wanted to perform helicase assays to determine if $DDX17_{135–555}$ can unwind RVFV ncRNA. Helicase assays are often conducted by using radioactivity or fluorescent resonance energy transfer (FRET) based analysis (53,54). However, our endeavor to develop a time and cost-effective alternative led us to design a unique experiment using MST. MST is ideal for our experiment because of its sensitivity for binding events, the low concentrations of samples required, and the availability of the reaction

components, other than fluorescently labeled DNA oligos (55,56). Using this simple assay, we demonstrated that $DDX17_{135–555}$ was able to unwind both RNAs and in a manner that is ATP dependent (Figure 7.7B). Currently, it is speculated that the ATP-binding domain hydrolyzes ATP to drive the helicase activity (57) which is consistent with our results. In conclusion, we have demonstrated that $DDX17_{135–555}$ is capable of directly binding and unwinding the non-coding regions of the S-segment genome of Rift Valley fever virus. This suggests that it could be critical for recognizing non-coding regions from other viral RNA.

7.7 Materials and Methods

7.7.1 Protein Expression and Purification of $DDX17_{135–555}$

The $DDX17_{135–555}$ cDNA construct in the pOPINF vector was designed with the help from the Oxford Protein Production Facility (OPPF, Harwell Oxford, Didcot, UK). $DDX17_{135–555}$ was expressed using Lemo21(DE3) *E. coli* cells. The culture was grown in Luria broth containing kanamycin (50 mg/mL) and chloramphenicol (100 mg/mL) anti-biotics. The culture was then transferred to Terrific broth containing 5% glycerol, and the cells were grown at 37 °C in an orbital shaker for 5 h, followed by a reduction in temperature to 20 °C for 16–18 h, harvested by centrifugation, and resuspended in lysis buffer (50 mM Tris, 500 mM NaCl, 10 mM imidazole 3 mM 2-mercaptoethanol, 10mg/mL lysozyme, 0.1% Tween-20, and 5% glycerol). The resulting cell suspension was sonicated and centrifuged at 30,000× g. The supernatant was filtered through a 0.45 µm syringe filter to prepare for chromatography.

Nickel affinity purification was performed using the ÄKTA start protein purification system (Global Life Science Solutions USA LLC, Marlborough, MA) with the HisTrap™ High-Performance column (Global Life Science Solutions USA LLC, Marlborough, MA) via the hexahistidine tag on $DDX17_{135–555}$. Protein was eluted in 2 mL fractions using an imidazole gradient up to 500 mM. Further purification and buffer exchange were performed using an ÄKTA pure purification system (Global Life Science Solutions USA LLC, Marlborough, MA) using

Superdex® 200 10/300 GL (Global Life Science Solutions USA LLC, Marlborough, MA). $DDX17_{135–555}$ was eluted in 50 mM Tris, 150 mM NaCl, and 3% glycerol. Elutions containing $DDX17_{135–555}$ were pooled and concentrated using Amicon® Ultra-15 Centrifugal Filter Units (30,000 MWCO, Millipore Canada Ltd, Etobicoke, ON). The 110μM $DDX17_{135–555}$ stocks were aliquoted and frozen in liquid nitrogen before being stored at −80 °C.

7.7.2 Preparation of Rift Valley Fever Virus Non-Coding RNAs

The cDNA sequences were prepared under T7 RNA polymerase control, with two additional G nucleotides on the 5’ end followed by an XbaI restriction enzyme cut site (T^CTAGA) on the 3’ end. Both RVRV constructs were designed based on the Genebank sequence of EU312119.1. The underlined regions are the complimentary regions to our fluorescent oligos described in a later section. Both RNA constructs used in the experiments are listed as follows:

1. RVFV NCR S Segment 812–886

5’GGAUUUGUUGAGGUUGAUUAGAGGUUAAGGCUGCCCCACCCCCCACCCCCUAAUCCCGACCGUAACCCCAACUCCU3’

2. RVFV IGR S Segment 25–100

5’GGCAAGUAUAUCAUGGAUUACUUUCCUGUGAUAUCUGUUGAUUUGCAGAGUGGUCGUCGUGUUGUGUCAGUGGAGUACAU3’

Each RNA was prepared using an in vitro transcription reaction using T7 RNA polymerase (made in-house) followed by purification using a Superdex® 200 10/300 GL via an ÄKTA pure system (Global Life Science Solutions USA LLC, Marlborough, MA). Fractions were analyzed using urea-polyacrylamide gel electrophoresis (urea-PAGE): 10 μL of each fraction were mixed with 2 μL of RNA loading dye and loaded into a 1.0 cm well PAGE (Bio-Rad Laboratories (Mississauga, ON). The urea-PAGE (7.5%) was then developed at 300 V,

room temperature for 25 min in 0.5× TBE, followed by staining and visualization with Sybr Safe (Thermofisher Scientific, Saint-Laurant, QC, Canada). Fractions containing a single band were used for further experimentation. Fractions containing the purified RNA of interest were concentrated by ethanol precipitation, and each pellet was resuspended in RNA buffer (10 mM Tris pH 7.5, 100 mM NaCl, and 5 mM $MgCl_2$).

7.7.3 Fluorescent Labeling of RNA

RNAs were incubated on ice for 30 min in 0.1M sodium acetate (pH 5.3) and 2 mM potassium periodate. Following incubation, the reaction was stopped by adding ethylene glycol to a concentration of 10 mM and incubated again on ice for 10 min. We then performed two ethanol precipitations, resuspended the RNA in water, along with 0.1 M NaOAc and 10 mM fluorescein-5-thiosemicarbazide (FITC), and incubated the mixture on ice and in the dark for 16 h. Following incubation with the fluorescent dye, the mixture was phenol extracted (1 vol phenol:1 vol mixture) five times until the phenol layer no longer changed color, indicating all free dye had been removed from the RNA mixture. We then ethanol precipitated the resulting labeled RNA twice, followed by re-suspension in RNA buffer.

7.7.4 Analytical Ultracentrifugation (AUC)

We collected SV-AUC data for FPLC-purified RNA and protein using a Beckman Optima AUC centrifuge with an AN60-Ti rotor at 20 °C. Each sample was loaded into Epon-2 channel centerpieces and was measured at 0.5 OD260 for RNA (680 nM) and 0.5 OD280 for protein (10.2 µM). For SV-AUC experiments, we used 10 mM Tris and 500 mM NaCl with 5 mM mgCl2 buffer at pH 7.5 for RNA and 50 mM Tris, 150 mM NaCl, and 5% glycerol buffer at pH 8 for protein. Intensity scans were collected at 20 s intervals at 40,000 revolutions per minute, measuring at 20 °C. All data were analyzed using UltraScan-III (18,58) according to the workflow described elsewhere (59). Finite element fits were processed on the Lonestar5 (Texas Advanced Computing Center, Austin, TX, USA) and Comet (San Diego Supercomputing

Center, San Diego, CA, USA) supercomputers. The collected SV-AUC data were analyzed using two-dimensional spectrum analysis (2DSA) to subtract time and radially invariant noise components and to fit the meniscus and bottom positions (60), followed by genetic algorithm analysis combined with Monte Carlo analysis (61). The buffer density and viscosity corrections were calculated with UltraScan (1.0030 g/cm3 and 1.0100 cP, respectively, for the RNA buffer and 1.017 g/cm3 and 1.152 cP for the protein buffer). Partial specific volumes of 0.55 mL/g [20] and 0.732 mL/g (62) were assumed for RNAs and protein, respectively. All reported hydrodynamic parameters are corrected to standard conditions (20 °C and water), as implemented in UltraScan (59).

7.7.5 Microscale Thermophoresis RNA and Protein Binding Studies

A two-fold serial dilution was performed on $DDX17_{135–555}$ where the highest concentration was 55 μM (as presented in Figure 7.6B). A constant amount of FITC-labeled RVFV NCR, or 5′ IGR, was added to each serial dilution of $DDX17_{135–555}$, resulting in a final concentration of 40 nM. The final concentration of polyU (negative control, Sigma-Aldrich Canada) in each assay was 50 μg/mL, and the initial fluorescence was similar to the ncRNA experiments. Samples were incubated together at room temperature for 10 min and then added to Nanotemper Technologies Monolith® NT.115 instrument (Munich, Germany) hydrophobic capillaries and loaded onto the MST block. Thermophoresis was measured at an ambient room temperature of 25 °C and performed using 20% excitation power for RVFV NCR and 40% for 5′ IGR (blue filter) and medium MST IR-laser power. Fluorescent migration used to determine K_d was measured from 4.0 to 5.0s and then normalized to initial fluorescence (−1.0 to 0s). The data from three independent replicates were analyzed using MO Affinity Analysis software v2.1.3 and fit to the standard K_d fit model, which describes a molecular interaction with a 1:1 stoichiometry according to the law of mass action. K_d is estimated by fitting Equation (1), where F(c) is the fraction bound at a given ligand concentration c;

Unbound is the F_{norm} signal of the target alone; Bound is the F_{norm} signal of the complex; K_d is the dissociation constant; and c_{target} is the final concentration of the target in the assay.

$$F(c) = Unbound + (Bound - Unbound) \times \frac{c + c_{target} + K_d - \sqrt{(c + c_{target} + K_d)^2 - 4cc_{target}}}{2c_{target}} \quad (7.1)$$

7.7.6 Helicase Assay

Firstly, we input our sequences into sfold (63) to determine the theoretical secondary structure and identified a portion of each RNA molecule that was double stranded. Oligos with complementary sequences to the double-stranded region(s) of the RNA(s) analyzed here were synthesized with a 5' conjugated Cy5 fluorophore. The region of each RNA molecule which the oligos hybridize to is underlined, as described above (4.3). The sequences for RVFV 5' IGR and RVFV 5' NCR oligo(s) are: 5'Cy5/CAACTCCAACTAATCTCCA3' and 5'Cy5/AGACAACTAAACGTCTCAC3', respectively.

Using Monolith® NT.115 that assesses the change in fluorescence migration, we were able to determine if the RNA molecules were unwound, thus allowing the oligo to bind to the now exposed complementary RNA. The reaction mixture contains 40nM of Cy5-DNA oligos, 1 μM of the RNA, and 4.25 mM of ATP. To test the helicase activity of $DDX17_{135–555}$, we added the enzyme to a final concentration of 20 μM. As a control, we compared the unwinding activity of bovine serum albumin (BSA) with the activity ob-served in the absence of any protein (black bar). For the BSA control, the same concentration was used as for DDX17 (green and blue bars). Additionally, to assess the im-portance of ATP in unwinding activity, we compared $DDX17_{135–555}$ without ATP to DDX17 with ATP (gray bar). Each run uses 4 capillaries, and we performed 3 runs for each condition before using the MO Affinity Analysis software to analyze the data. The analysis software assesses the signal to noise ratio between a run with and without the protein. Signal to noise is a measure of the response amplitude

that is divided by the noise of the environment, and Equation (2) represents how this can be calculated (64). If the signal to noise ratio rises above 5, the assay indicates that a binding event has occurred. To further analyze the helicase assay, unpaired t-tests were performed.

$$\frac{S}{N} = \frac{ResponseAmplitude}{\sqrt{\frac{\sum_i (r_i - \acute{r})^2}{n-1}}} \tag{7.2}$$

7.7.7 Small-Angle X-ray Scattering

Small-angle X-ray scattering was performed by utilizing the B21 BioSAXS beamline at Diamond Light Source (Didcot, Oxfordshire, UK) to collect high-performance liquid chromatography SAXS (HPLC-SAXS) data which can be found described previously [66]. Using a specialized flow cell connected to an in-line Agilent 1200 (Agilent Technologies, Stockport, UK) HPLC, 50µL of each purified sample (protein or RNA) were injected onto a Shodex KW403-4F (Showa Denko America Inc., New York, NY, USA) size exclusion column pre-equilibrated with buffer, at a flow rate of 0.160 mL per minute. X-rays were exposed to each frame for 3 s. The peak region for each sample was buffer subtracted using baseline measurements and merged using Primus (65) or ScAtter (66), as previously described. The merged data were analyzed initially by Guinier approximation (22) to obtain the radius of gyration (R_g) and evaluate homogeneity. Dimensionless Kratky analysis (24) was performed on all samples to evaluate the folding extend of the biomolecules of interest, which is reviewed in detail elsewhere (23). Following Kratky analysis, we performed a pair distance distribution (P(r)) analysis using GNOM (25) to additionally provide the R_g and the maximum particle dimension (D_{max}). Using the information from the P(r) plot, we generated models using DAMMIN (27), without enforced symmetry, which can be found previously described (30). Finally, the resulting models were averaged and filtered to generate a single representative averaged model using DAMAVER (28,43,67).

Recently, a crystal structure of DDX17 containing the ATP and helicase domain (6UV0) was published (29). We used the scattering data of $DDX17_{135–555}$ and performed high-resolution modeling, using the crystal structure and CORAL program, as described earlier (30). Briefly, the high-resolution structure information of the ATP domain (155aa–382) and helicase domain (389aa–555) was provided as input data along with the raw scattering data, and the residues 383–388 were used as a flexible linker. Using this approach, we initially calculated 12 models and the quality of the models was assessed using χ^2 values.

7.8 References

1. Findlay, G. and Daubney, R. (1931) The Virus of Rift Valley Fever or Enzob'tic Hepatitis. *Lancet*.
2. Balkhy, H.H. and Memish, Z.A. (2003) Rift Valley fever: an uninvited zoonosis in the Arabian peninsula. *International journal of antimicrobial agents*, **21**, 153-157.
3. Bird, B.H., Ksiazek, T.G., Nichol, S.T. and MacLachlan, N.J. (2009) Rift Valley fever virus. *Journal of the American Veterinary Medical Association*, **234**, 883-893.
4. Ikegami, T. (2017) Rift Valley fever vaccines: an overview of the safety and efficacy of the live-attenuated MP-12 vaccine candidate. *Expert review of vaccines*, **16**, 601-611.
5. WHO. (2020) Rift Valley Fever Virus.
6. Linthicum, K., Davies, F., Kairo, A. and Bailey, C. (1985) Rift Valley fever virus (family Bunyaviridae, genus Phlebovirus). Isolations from Diptera collected during an inter-epizootic period in Kenya. *Epidemiology & Infection*, **95**, 197-209.
7. Nanyingi, M.O., Munyua, P., Kiama, S.G., Muchemi, G.M., Thumbi, S.M., Bitek, A.O., Bett, B., Muriithi, R.M. and Njenga, M.K. (2015) A systematic review of Rift Valley Fever epidemiology 1931–2014. *Infection ecology & epidemiology*, **5**, 28024.
8. Terasaki, K., Murakami, S., Lokugamage, K.G. and Makino, S. (2011) Mechanism of tripartite RNA genome packaging in Rift Valley fever virus. *Proceedings of the National Academy of Sciences*, **108**, 804-809.
9. Pepin, M., Bouloy, M., Bird, B.H., Kemp, A. and Paweska, J. (2010) Rift Valley fever virus (Bunyaviridae: Phlebovirus): an update on pathogenesis, molecular epidemiology, vectors, diagnostics and prevention. *Veterinary research*, **41**.
10. Ikegami, T. and Makino, S. (2011) The pathogenesis of Rift Valley fever. *Viruses*, **3**, 493-519.
11. Moy, R.H., Cole, B.S., Yasunaga, A., Gold, B., Shankarling, G., Varble, A., Molleston, J.M., Lynch, K.W. and Cherry, S. (2014) Stem-loop recognition by DDX17 facilitates miRNA processing and antiviral defense. *Cell*, **158**, 764-777.
12. Gauliard, N., Billecocq, A., Flick, R. and Bouloy, M. (2006) Rift Valley fever virus noncoding regions of L, M and S segments regulate RNA synthesis. *Virology*, **351**, 170-179.
13. Byrd, A.K. and Raney, K.D. (2012) Superfamily 2 helicases. *Frontiers in bioscience (Landmark edition)*, **17**, 2070.
14. Meier-Stephenson, V., Mrozowich, T., Pham, M. and Patel, T.R. (2018) DEAD-box helicases: the Yin and Yang roles in viral infections. *Biotechnology and Genetic Engineering Reviews*, **34**, 3-32.
15. Linder, P. and Jankowsky, E. (2011) From unwinding to clamping—the DEAD box RNA helicase family. *Nature reviews Molecular cell biology*, **12**, 505-516.
16. Fuller-Pace, F.V. (2013) The DEAD box proteins DDX5 (p68) and DDX17 (p72): multi-tasking transcriptional regulators. *Biochimica Et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, **1829**, 756-763.
17. Rozen, F., Edery, I., Meerovitch, K., Dever, T.E., Merrick, W.C. and Sonenberg, N. (1990) Bidirectional RNA helicase activity of eucaryotic translation initiation factors 4A and 4F. *Molecular and cellular biology*, **10**, 1134-1144.
18. Demeler, B. (2005) UltraScan: a comprehensive data analysis software package for analytical ultracentrifugation experiments. *Modern analytical ultracentrifugation: techniques and methods*, **10**, 210-229.
19. Brosey, C.A. and Tainer, J.A. (2019) Evolving SAXS versatility: solution X-ray scattering for macromolecular architecture, functional landscapes, and integrative structural biology. *Current opinion in structural biology*, **58**, 197-213.

20. Pérez, J. and Vachette, P. (2017) A successful combination: coupling SE-HPLC with SAXS. *Biological small angle scattering: techniques, strategies and tips*, 183-199.
21. Mrozowich, T., Henrickson, A., Demeler, B. and Patel, T.R. (2020) Nanoscale structure determination of murray valley encephalitis and powassan virus non-coding RNAs. *Viruses*, **12**, 190.
22. Guinier, A., Fournet, G. and Yudowitch, K.L. (1955) Small-angle scattering of X-rays.
23. Patel, T.R., Chojnowski, G., Koul, A., McKenna, S.A. and Bujnicki, J.M. (2017) Structural studies of RNA-protein complexes: A hybrid approach involving hydrodynamics, scattering, and computational methods. *Methods*, **118**, 146-162.
24. Durand, D., Vivès, C., Cannella, D., Pérez, J., Pebay-Peyroula, E., Vachette, P. and Fieschi, F. (2010) NADPH oxidase activator p67phox behaves in solution as a multidomain protein with semi-flexible linkers. *Journal of structural biology*, **169**, 45-53.
25. Svergun, D. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of applied crystallography*, **25**, 495-503.
26. Svergun, D.I. and Koch, M.H. (2003) Small-angle scattering studies of biological macromolecules in solution. *Reports on Progress in Physics*, **66**, 1735.
27. Svergun, D.I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophysical journal*, **76**, 2879-2886.
28. Volkov, V.V. and Svergun, D.I. (2003) Uniqueness of ab initio shape determination in small-angle scattering. *Journal of applied crystallography*, **36**, 860-864.
29. Ngo, T.D., Partin, A.C. and Nam, Y. (2019) RNA specificity and autoregulation of DDX17, a modulator of microRNA biogenesis. *Cell reports*, **29**, 4024-4035. e4025.
30. Reuten, R., Patel, T.R., McDougall, M., Rama, N., Nikodemus, D., Gibert, B., Delcros, J.-G., Prein, C., Meier, M. and Metzger, S. (2016) Structural decoding of netrin-4 reveals a regulatory function towards mature basement membranes. *Nature communications*, **7**, 13515.
31. Mrozowich, T., MeierStephenson, V. and Patel, T.R. (2019) Microscale thermophoresis: warming up to a new biomolecular interaction technique. *The Biochemist*, **41**, 8-12.
32. Jerabek-Willemsen, M., Wienken, C.J., Braun, D., Baaske, P. and Duhr, S. (2011) Molecular interaction studies using microscale thermophoresis. *Assay and drug development technologies*, **9**, 342-353.
33. Jerabek-Willemsen, M., André, T., Wanner, R., Roth, H.M., Duhr, S., Baaske, P. and Breitsprecher, D. (2014) MicroScale Thermophoresis: Interaction analysis and beyond. *Journal of Molecular Structure*, **1077**, 101-113.
34. Schuck, P., Perugini, M.A., Gonzales, N.R., Howlett, G.J. and Schubert, D. (2002) Size-distribution analysis of proteins by analytical ultracentrifugation: strategies and application to model systems. *Biophysical journal*, **82**, 1096-1111.
35. Chillón, I., Marcia, M., Legiewicz, M., Liu, F., Somarowthu, S. and Pyle, A.M. (2015), *Methods in enzymology*. Elsevier, Vol. 558, pp. 3-37.
36. Patel, T.R., Winzor, D.J. and Scott, D.J. (2016) Analytical ultracentrifugation: A versatile tool for the characterisation of macromolecular complexes in solution. *Methods*, **95**, 55-61.
37. Chen, Y. and Pollack, L. (2016) SAXS studies of RNA: structures, dynamics, and interactions with partners. *Wiley Interdisciplinary Reviews: RNA*, **7**, 512-526.
38. Kim, D.N., Thiel, B.C., Mrozowich, T., Hennelly, S.P., Hofacker, I.L., Patel, T.R. and Sanbonmatsu, K.Y. (2020) Zinc-finger protein CNBP alters the 3-D structure of lncRNA Braveheart in solution. *Nature communications*, **11**, 148.
39. Patel, T.R., Meier, M., Li, J., Morris, G., Rowe, A.J. and Stetefeld, J. (2011) T-shaped arrangement of the recombinant agrin G3–IgG Fc protein. *Protein Science*, **20**, 931-940.

40. Krahn, N., Meier, M., To, V., Booy, E.P., McEleney, K., O'Neil, J.D., McKenna, S.A., Patel, T.R. and Stetefeld, J. (2017) Nanoscale assembly of high-mobility group AT-Hook 2 protein with DNA replication fork. *Biophysical journal*, **113**, 2609-2620.
41. Deo, S., Patel, T.R., Chojnowski, G., Koul, A., Dzananovic, E., McEleney, K., Bujnicki, J.M. and McKenna, S.A. (2015) Characterization of the termini of the West Nile virus genome and their interactions with the small isoform of the 2' 5'-oligoadenylate synthetase family. *Journal of structural biology*, **190**, 236-249.
42. Dzananovic, E., Chojnowski, G., Deo, S., Booy, E.P., Padilla-Meier, P., McEleney, K., Bujnicki, J.M., Patel, T.R. and McKenna, S.A. (2017) Impact of the structural integrity of the three-way junction of adenovirus VAI RNA on PKR inhibition. *PloS one*, **12**, e0186849.
43. Dzananovic, E., Patel, T.R., Chojnowski, G., Boniecki, M.J., Deo, S., McEleney, K., Harding, S.E., Bujnicki, J.M. and McKenna, S.A. (2014) Solution conformation of adenovirus virus associated RNA-I and its interaction with PKR. *Journal of structural biology*, **185**, 48-57.
44. Burchard, W. (1994) Light scattering techniques. *Physical techniques for the study of food biopolymers*, 151-213.
45. Stetefeld, J., McKenna, S.A. and Patel, T.R. (2016) Dynamic light scattering: a practical guide and applications in biomedical sciences. *Biophysical reviews*, **8**, 409-427.
46. Ferens, F.G., Patel, T.R., Oriss, G. and Stetefeld, J. (2019) A cholesterol analog induces an oligomeric reorganization of VDAC. *Biophysical journal*, **116**, 847-859.
47. Moon, M.H., Hilimire, T.A., Sanders, A.M. and Schneekloth Jr, J.S. (2018) Measuring RNA–ligand interactions with microscale thermophoresis. *Biochemistry*, **57**, 4638-4643.
48. Xing, Z., Wang, S. and Tran, E.J. (2017) Characterization of the mammalian DEAD-box protein DDX5 reveals functional conservation with S. cerevisiae ortholog Dbp2 in transcriptional control and glucose metabolism. *RNA*, **23**, 1125-1138.
49. Wu, G., Xing, Z., Tran, E.J. and Yang, D. (2019) DDX5 helicase resolves G-quadruplex and is involved in *MYC* gene transcriptional activation. *Proceedings of the National Academy of Sciences*, **116**, 20453-20461.
50. Moy, R.H. and Cherry, S. (2014) DDX17: Structured RNA recognition drives diverse outputs. *Cell Cycle*, **13**, 3467-3468.
51. Dzananovic, E., Patel, T.R., Deo, S., McEleney, K., Stetefeld, J. and McKenna, S.A. (2013) Recognition of viral RNA stem–loops by the tandem double-stranded RNA binding domains of PKR. *Rna*, **19**, 333-344.
52. Song, H. and Ji, X. (2019) The mechanism of RNA duplex recognition and unwinding by DEAD-box helicase DDX3X. *Nature communications*, **10**, 3085.
53. Mendoza, O., Gueddouda, N.M., Boulé, J.-B., Bourdoncle, A. and Mergny, J.-L. (2015) A fluorescence-based helicase assay: application to the screening of G-quadruplex ligands. *Nucleic acids research*, **43**, e71-e71.
54. Mojumdar, A. and Deka, J. (2019) Assaying the activity of helicases: an overview. *Helicases from All Domains of Life*, 235-246.
55. Cordin, O., Tanner, N.K., Doere, M., Linder, P. and Banroques, J. (2004) The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity. *EMBO J*, **23**, 2478-2487.
56. Tani, H., Akimitsu, N., Fujita, O., Matsuda, Y., Miyata, R., Tsuneda, S., Igarashi, M., Sekiguchi, Y. and Noda, N. (2009) High-throughput screening assay of hepatitis C virus helicase inhibitors using fluorescence-quenching phenomenon. *Biochemical and biophysical research communications*, **379**, 1054-1059.
57. Lamm, G.M., Nicol, S.M., Fuller-Pace, F.V. and Lamond, A.I. (1996) p72: a human nuclear DEAD box protein highly related to p68. *Nucleic acids research*, **24**, 3739-3747.

58. Demeler, B. (2010) Methods for the design and analysis of sedimentation velocity and sedimentation equilibrium experiments with proteins. *Current protocols in protein science*, **60**, 7.13. 11-17.13. 24.
59. Demeler, B. and Gorbet, G.E. (2016) Analytical ultracentrifugation data analysis with UltraScan-III. *Analytical ultracentrifugation: instrumentation, software, and applications*, 119-143.
60. Brookes, E., Cao, W. and Demeler, B. (2010) A two-dimensional spectrum analysis for sedimentation velocity experiments of mixtures with heterogeneity in molecular weight and shape. *European Biophysics Journal*, **39**, 405-414.
61. Brookes, E.H. and Demeler, B. (2007), *Proceedings of the 9th annual conference on Genetic and evolutionary computation*, pp. 361-368.
62. Durchschlag, H. (2021), *Thermodynamic data for biochemistry and biotechnology*. Springer, pp. 45-128.
63. Ding, Y., Chan, C.Y. and Lawrence, C.E. (2004) S fold web server for statistical folding and rational design of nucleic acids. *Nucleic acids research*, **32**, W135-W141.
64. Seidel, S.A., Dijkman, P.M., Lea, W.A., van den Bogaart, G., Jerabek-Willemsen, M., Lazic, A., Joseph, J.S., Srinivasan, P., Baaske, P. and Simeonov, A. (2013) Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. *Methods*, **59**, 301-315.
65. Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H. and Svergun, D.I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of applied crystallography*, **36**, 1277-1282.
66. Rambo, R.P. (2017) ScÅtter, a JAVA-based application for basic analysis of SAXS datasets.
67. Patel, T.R., Bernards, C., Meier, M., McEleney, K., Winzor, D.J., Koch, M. and Stetefeld, J. (2014) Structural elucidation of full-length nidogen and the laminin–nidogen complex in solution. *Matrix Biology*, **33**, 60-67.

Chapter 8. Human DDX3X Unwinds Japanese Encephalitis and Zika Viral 5′ Terminal Regions

8.1 Foreword

This chapter consists of a manuscript I wrote in collaboration with Corey R Nelson (MSc 2021), whom I share co-authorship with, Sean Park, Darren Gemmill, and Dr. Trushar R Patel, published in *The International Journal of Molecular Sciences,* Volume 22, issue 1 on January 2, 2021. https://doi.org/10.3390/ijms22010413. This publication focuses on the interaction of JEV and Zika 5′ non-coding RNA terminal regions with DDX3X human RNA helicase. Corey and I were both involved in all the experiments manuscript with him focusing more on the protein aspect, and I the RNA aspect. Darren Gemmill provided the Zika 5′ RNA. We both equally contributed to the writing of the final draft. This publication is reproduced with permission from Corey Nelson, MDPI, and the Creative Commons CC by license and re-formatted to fit thesis formatting.

8.2. Author List and Affiliations

Corey Nelson [1†,] **Tyler Mrozowich** [1†,] Darren L. Gemmill [1], Sean M. Park [1] and Trushar R. Patel [1,2,3]

1. Department of Chemistry and Biochemistry, Alberta RNA Research and Training Institute, University of Lethbridge, 4401 University Drive, Lethbridge, AB T1K 3M4, Canada
2. Department of Chemistry and Biochemistry, University of Montana, Missoula, MT 59812, USA
3. Li Ka Shing Institute of Virology and Discovery Lab, University of Alberta, Edmonton, AB T6G 2E1, Canada

[†]Equal Author Contribution

8.3 Abstract

Flavivirus genus includes many deadly viruses such as the Japanese encephalitis virus and Zika virus. The 5′ terminal regions of flaviviruses interact with human proteins and such interactions are critical for viral replication. One of the human proteins identified to interact with the 5′ TR of JEV is the DEAD-box helicase, DDX3X. In this study, we *in vitro* transcribed the 5′ TR of JEV and demonstrated its direct interaction with recombinant DDX3X (K_d of 1.66

± 0.21 μM) using microscale thermophoresis (MST). Due to the proposed structural similarities of 5′ and 3′ TRs of flaviviruses, we investigated if the ZIKV 5′ TR could also interact with human DDX3X. Our MST studies suggested that DDX3X recognizes ZIKV 5′ TR with a K_d of 7.05 ± 0.75 μM. Next, we performed helicase assays that suggested that the binding of DDX3X leads to the unwinding of JEV and ZIKV 5′ TRs. Overall, our data indicate, for the first time, that DDX3X can directly bind and unwind in vitro transcribed flaviviral TRs. In summary, our work indicates that DDX3X could be further explored as a therapeutic target to inhibit Flaviviral replication

8.4 Introduction

Infection with pathogenic viruses often leads to severe diseases that may impact, among others, the metabolic, respiratory, digestive, and central nervous systems. Among the most pathogenic family of viruses, emerging and re-emerging outbreaks of flavivirus are responsible for thousands of deaths annually (1). Furthermore, flaviviral infections also lead to significant morbidities in survivors, which can create a substantial burden on the health system (2,3). The Flaviviridae family includes the most prevalent arthropod-borne viruses such as dengue, Japanese encephalitis (JEV), Murrey Valley, Powassan, West Nile, yellow fever, and Zika viruses. Flaviviral out-breaks are becoming increasingly common, due to the ease of transmission by mosquitoes and the lack of efficient therapeutics or immunoprophylactic strategies (4-7). As a result, flaviviruses are emerging as a global health threat. For example, the WHO reports that since 2015, ZIKV outbreaks have been reported in 42 countries (6). The majority of flaviviruses utilize Aedes and Culex genus' of mosquitoes for transmission (8), meaning that as global temperatures increase, countries that were once protected from arboviruses are also becoming increasingly at risk (9,10).

The JEV is responsible for approximately 68,000 cases annually, with a fatality rate between 20–30% and with 30% of cases developing serious long-term disabilities (11,12). This makes JEV one of the deadliest flaviviruses, although an approved vaccine against JEV

is available (13,14). The Zika virus outbreak infected >4.5 million people in Brazil and the Americas in 2015–2016 (15,16). ZIKV is also linked to birth defects (2,17-23) and neurological disorders (24-30). The ZIKV has also been observed to be trans-mitted sexually and present symptoms in only one-fifth of cases (31,32). Unlike JEV, there is no approved vaccine available against ZIKV. Attempts to develop vaccines have faced unexpected challenges due to antibody-dependent enhancement of infection with other flaviviruses (33-35). Flaviviruses contain a positive-sense single-stranded RNA genome, comprised of 5′ and 3′ untranslated terminal regions (TRs). The 5′ TRs (~0.1 kb) contain a type 1 capped structure, whereas the 3′ TRs (~0.3 to ~0.5 kb) lack a poly(A) tail, and both TRs include conserved structural motifs (36-42). The interactions between flaviviral 5′ and 3′ TRs are also critical for viral replication (4,42-47). These regions flank a single open reading frame (ORF), which encodes a single polypeptide that is cleaved by a combination of host and viral proteases (48). Previous work has established that interactions of host proteins with the flaviviral TRs are crucial for viral replication (49-60).

The human DEAD-box family of helicases is comprised of 37 members, each composed of conserved helicase core domains that interact with ATP and RNA (61,62). The DEAD-box helicases consist of two helicase domains; domain 1 containing motifs Q, I, Ia Ib, Ic, II (DEAD-box), and III; and domain 2 containing motifs IV, IVa, V, Va, and VI (62). These motifs are involved in either ATP binding/hydrolysis, RNA binding, or couple ATP and RNA binding activities. Although the classical function of DEAD-box helicases is ATP-dependent unwinding of nucleic acids (63), they influence all major aspects of RNA metabolism (64). DDX3X (X-linked DDX3, 73 kDa, Figure 8.1A) is one of the ATP-dependent RNA helicases that plays critical roles in transcription, translation, and mRNA export (60,65,66). DDX3X unwinds RNA in an ATP-dependent manner, where DDX3X binds to dsRNA (double-stranded RNA) and hydrolyzes ATP to release single-stranded RNA (67). Apart from playing vital roles in cellular activities, DDX3X was shown to suppress dengue viral infection via interferon activation (68)

but promotes WNV infection (69). Using affinity pull-down and Western blot analysis, it was suggested that DDX3X could interact with 5′ TR of JEV and regulate its replication (51). Recent studies have suggested that DDX3X inhibitors can suppress WNV replication (69). Given the magnitude and severity of flaviviral infections, there is a critical need for therapeutics; however, their development is hindered by the limited understanding of the interactions of viral RNAs with the host cellular proteins.

In this study, we followed up Li et al.'s affinity pull-down assays and demonstrated that DDX3X directly interacts with the 5′ TR of JEV [50]. As the 5′ TRs of flaviviruses are hypothesized to be structurally similar (70,71), we asked if the 5′ TR of ZIKV can also be recognized by DDX3X. Our binding studies demonstrated that DDX3X indeed interacts with the 5′ TR of ZIKV. As DDX3X is an RNA helicase, we performed helicase assays, which suggested that both JEV and ZIKV 5′ TR can be unwound by DDX3X. In summary, our study highlights that DDX3X could serve as an important therapeutic target to inhibit JEV and ZIKV replications.

8.5 Results

8.5.1 Purification of DDX3X$_{132–607}$, JEV 5′ and Zika 5′ TR RNAs

The DDX3X$_{132–607}$ was expressed in Lemo21(DE3) *Esherichia coli* cells, followed by initial purification using affinity chromatography. Subsequently, affinity-purified DDX3X$_{132–607}$ was subjected to size exclusion chromatography (SEC) purification (see section 4 for additional details). As presented in Figure 8.1B, we were able to remove minor aggregation at ~12 mL and ~13 mL elution volumes to obtain homogenous preparation of DDX3X$_{132–607}$ (peak at ~14.5 to 15.5 mL). Subsequently, we analysed the peak fractions using SDS-PAGE, which suggested that DDX3X$_{132–607}$ is devoid of any degradation (Figure 8.1C, right lane) and the amino-acid sequence-based molecular weight of 55.3 kDa for DDX3X$_{132–607}$ coincides with the observed band in Figure 8.1C.

The predicted secondary structure of 5′ TRs of JEV and ZIKA (Figure 8.2A,B) indicated that both RNAs are composed of a significant amount of double-stranded regions, along with stem-loops. The 5′ TRs for both viruses were in vitro transcribed and natively purified using SEC, similarly to DDX3X$_{132–607}$. The SEC purification indicated that while the ZIKV 5′ TR RNA elutes at approximately 12.5 mL, the JEV 5′ TR elutes at ~13.8 mL (Figure 8.2C). In both cases, oligomeric or aggregated species appear to elute around ~10.0 to 11.5 mL. In both SEC profiles, the plasmid DNA that was used as a template elutes around 8 mL, consistent with the column's void volume. Urea-PAGEs confirmed that both RNAs were purified to homogeneity. Monodispersed peak fractions(s) were used in downstream experiments.

8.5.2 DDX3X$_{132–607}$ *Binds to 5′ TRs of JEV and ZIKV*

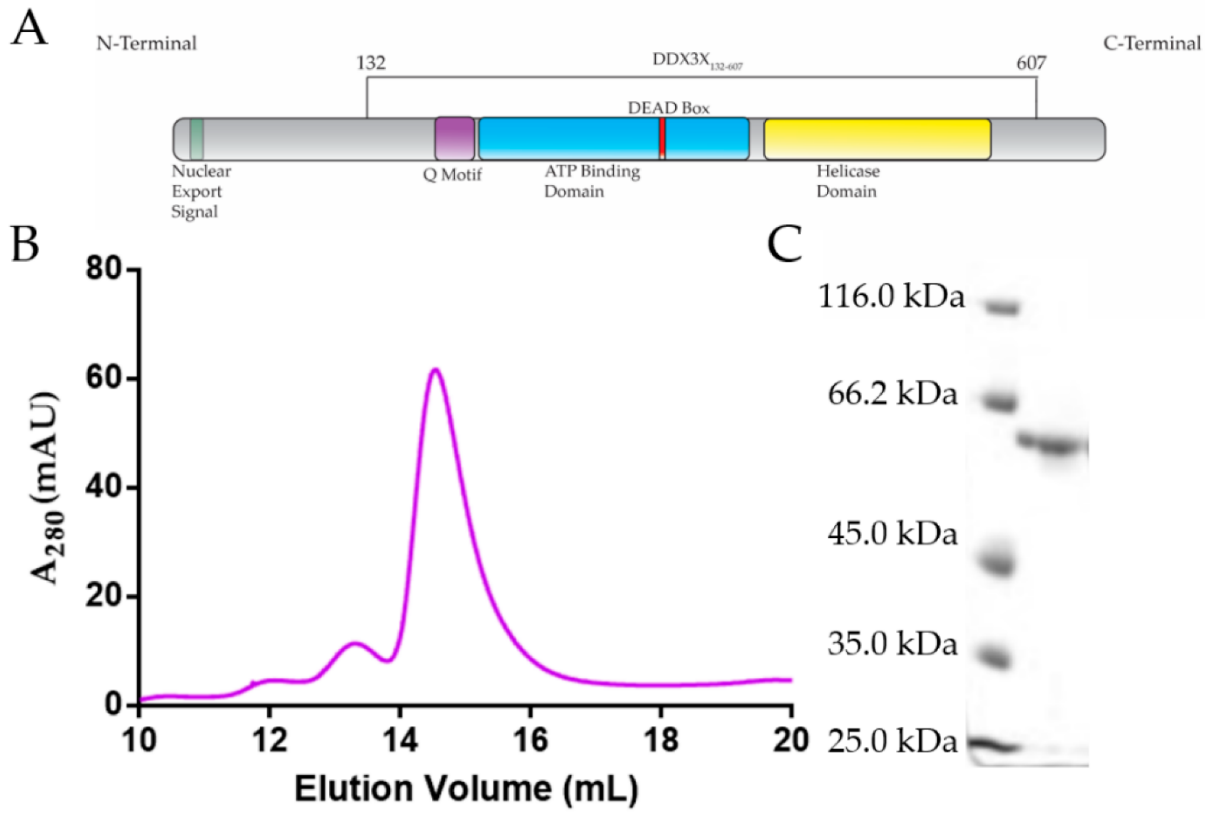


Figure 8.1. Purification of recombinant DDX3X$_{132–607}$. (A) Schematic of human DDX3X's domain architecture, indicating that DDX3X$_{132–607}$ consists of all the major domains, except the nuclear export signal sequence. **(B)** Size exclusion chromatography purification (Superdex 200 Increase GL 10/300) of DDX3X$_{132–607}$ demonstrating that DDX3X$_{132–607}$ can be purified to homogeneity, eluting at ~14.5 mL. Y-axis represents absorbance at 280 nm while the x-axis represents elution volume. **(C)** SDS-PAGE indicating that the size exclusion chromatography (SEC)-purified DDX3X132–607 is monodispersed with the correct molecular weight (55.3 kDa).

To determine the binding affinity of a previously uncharacterized interaction system containing $DDX3X_{132-607}$ and 5′ TR of JEV, we employed microscale thermophoresis as performed previously (72,73). $DDX3X_{132-607}$ was titrated against the constant concentration of the fluorescent ncRNAs (non-coding RNAs) to determine their dissociation constant (K_d). Figure 8.3A displays the change in fluorescent migration when the infrared laser affects the samples, where each trace represents a different concentration of $DDX3X_{132-607}$. Additionally, the blue bar indicates the "cold" region, and the red bar indicates the "hot" region. Figure

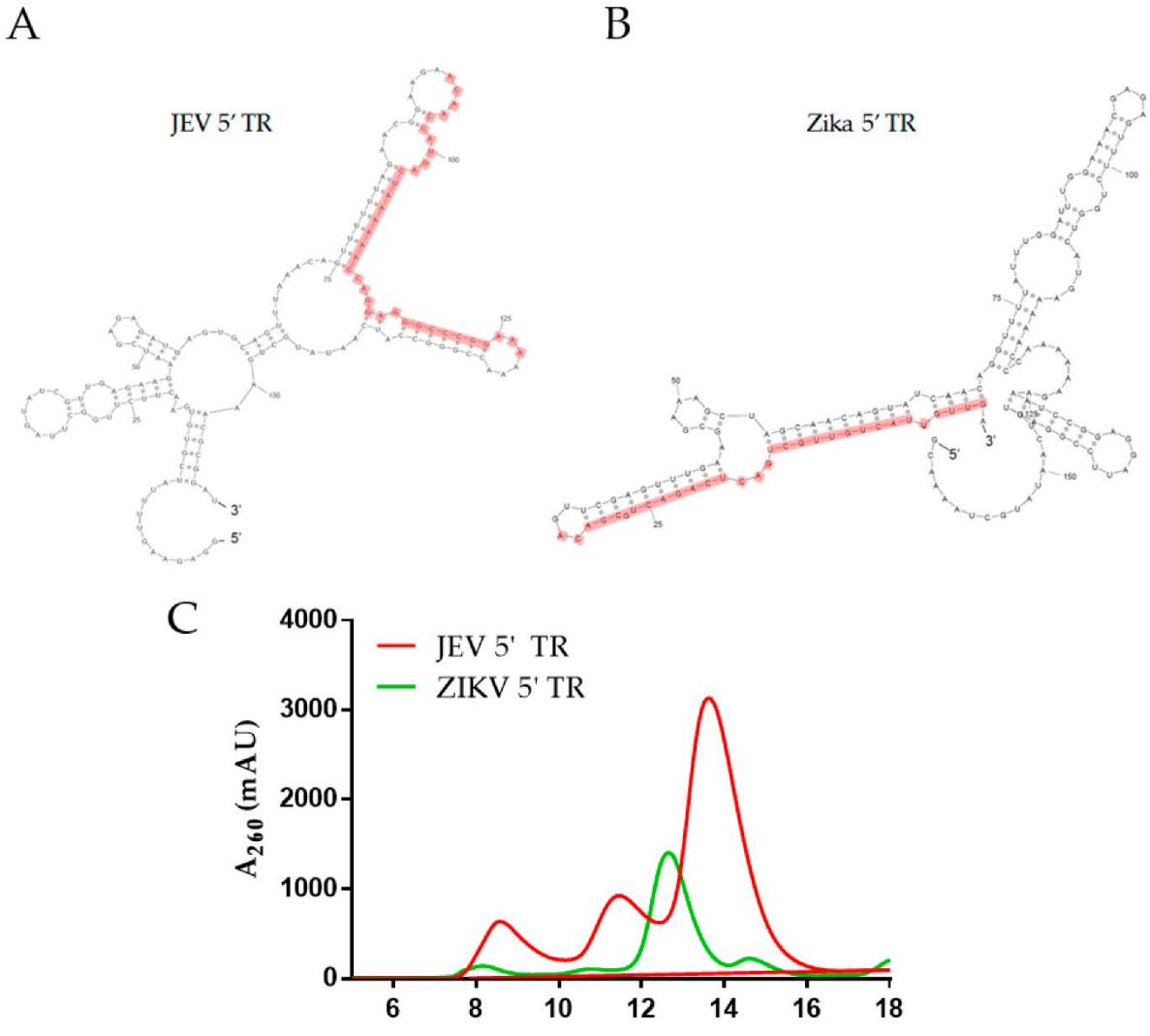


Figure 8.2. Purification of viral terminal regions. (A and B) Predicted secondary structure (sfold) for the 5′ terminal regions (TRs) of Japanese encephalitis virus (JEV) 5′ and Zika virus (ZIKV), respectively. The regions highlighted in red colour indicate the sequence complementary to the DNA oligo used in the helicase assays. (C) Size exclusion chromatography elution profiles of the 5′ TRs of JEV (red) and ZIKV (green). Arrows indicate peaks that represent monodispersed fractions of RNA, which were used for downstream experiments. Y-axis represents absorbance at 260 nm while the x-axis represents elution volume.

8.3B and 3C present the binding curves for 5′ TRs of JEV and ZIKV in red and green, respectively. These experiments suggest that DDX3X$_{132–607}$ has a K_d of 1.66 ± 0.2 1μM for the 5′ TR of JEV and 7.05 ± 0.75 μM for the 5′ TR of ZIKV.

8.5.3. DDX3X$_{132–607}$ *Unwinds 5′ TRs of JEV and ZIKV*

To further investigate if the binding of flaviviral 5′ TRs with DDX3X$_{132–607}$ leads to their unwinding, we designed a helicase assay using MST. We hypothesized that if DDX3X$_{132–607}$ unwinds the 5′ TRs of JEV and ZIKV, the newly formed single-stranded region of viral RNAs could hybridize with their complementary DNA oligos. Therefore, we designed fluorescently labelled DNA oligos complementary to the region highlighted in red colour in Figures 2A and B for JEV and ZIKV, respectively. We incubated 5′ TR of JEV (or ZIKV) with DDX3X$_{132–607}$ (or BSA as a negative control), ATP and complementary oligos, followed by measurements of MST traces. Our MST experiments suggest that in the presence of DDX3X$_{132–607}$, the oligos migrated differently than in the presence of BSA. This indicates that DDX3X$_{132–607}$ is able to unwind the double-stranded region, allowing the oligo to bind. Figure 8.4A and B present the unwinding of 5′ TRs of JEV and ZIKV with a signal to noise ratio of 14.8 ± 4.82 and 20.0 ± 5.06, respectively. Note that the signal to noise ratio of 12 or higher indicates the excellent

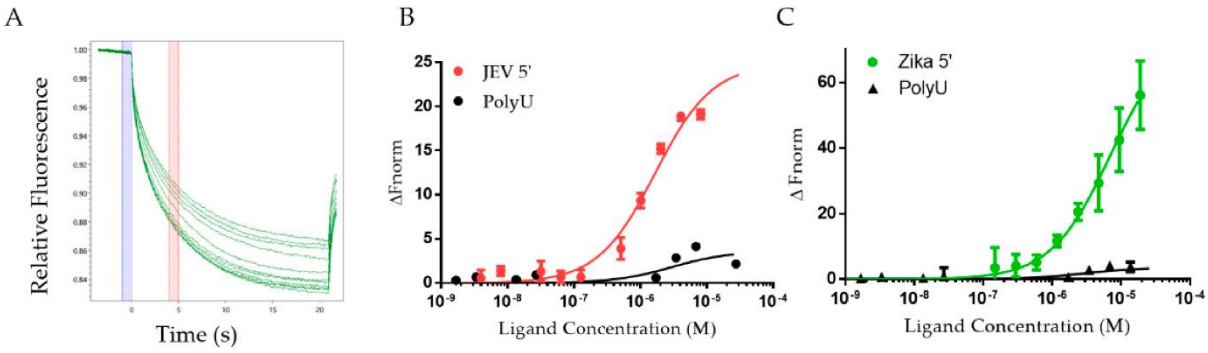


Figure 8.3. Interaction studies using microscale thermophoresis. (A) Representative MST traces depicting the change in fluorescent migration of fluorescein-5-thiosemicarbazide (FITC)-JEV 5′ TR due to the excitation with an infra-red laser. Each green trace correlates to a different concentration of DDX3X$_{132–607}$. (B) Binding data for DDX3X$_{132–607}$ with JEV 5′ TR (n = 3). The red curve represents JEV 5′ and has a dissociation constant of 1.66 ± 0.21μM (Std. error of regression = 1.25). The black trace represents that polyU (negative control) does not interact with DDX3X$_{132–607}$. (C) Interaction between DDX3X$_{132–607}$ and ZIKV 5′ TR (n = 3) studied using MST. The green curve represents ZIKV 5′ and has a dissociation constant of 7.05 ± 0.75μM (Std. error of regression = 1.22). PolyU binding data included for reference (black).

quality of the assay (74). To achieve the effect of a null mutant DDX3X, which can involve mutating the "DEAD" region in the ATP-binding domain (75), we compared the ability of $DDX3X_{132–607}$ to unwind ncRNA in the presence and absence of ATP. Figure 8.4C shows the change in fluorescent migration as a result of ATP. Adding ATP caused a significant change in the migration of the DNA oligos, indicating that the unwinding activity increased in the presence of ATP.

8.6 Discussion

The purification of $DDX3X_{132–607}$ (55.3 kDa) through SEC resulted in the protein eluting at ~14.5 mL (Figure 8.1B), which is consistent with our previous purification of a similar molecular weight human helicase, $DDX3X_{132–607}$ (48.5 kDa) (76). As both helicases are highly similar and belong to the same family of proteins, we are confident that the peak we observe for $DDX3X_{132–607}$ is consistent with a homogenous, monodispersed preparation. Further quality control was performed through SDS-PAGE, which con-firms that the peak SEC fractions contain a singular monodispersed species at the correct molecular weight, and are devoid of any degradation (Figure 8.1C). The sfold predicted secondary structures reveal that each RNA adopts a high degree of double-stranded regions (Figure 8.2A and 2B), which is consistent with previous reports of highly structured flaviviral terminal regions (71).

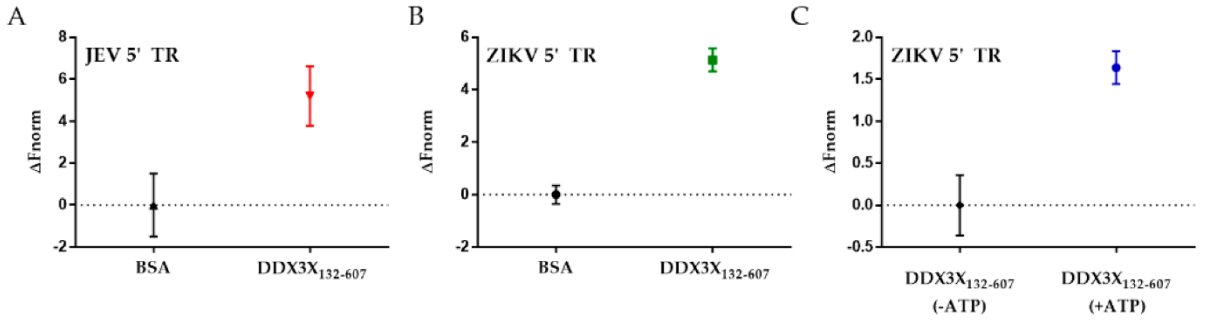


Figure 8.4. Helicase assays conducted using microscale thermophoresis. (A) Comparing the change of fluorescent migration of a complementary DNA oligo in the presence of 5′ TR of JEV with ATP and either bovine serum albumin (BSA) or $DDX3X_{132–607}$. The average signal to noise ratio is 14.8 ± 4.82 (n = 3). **(B)** Comparing the fluorescent migration of a complementary DNA oligo in the presence of ZIKV 5′ TR with ATP and either BSA or $DDX3X_{132–607}$. The average signal to noise is 20.0 ± 5.06 (n = 3). **(C)** Investigating the role of ATP in $_{DDX3X132–607}$ helicase activity. The average signal to noise ratio is 5.5 ± 0.59 (n = 3), suggesting that ATP is required to unwind RNA.

RNA purification was performed immediately after in vitro transcription reaction using SEC. The 5′ TRs of JEV (159 nts) and Zika (163 nts) have sequence molecular weights of 51.5 and 52.3 kDa, respectively. However, they elute at distinctly different positions during SEC purification (Figure 8.1C). We believe that the differences in their structures lead to their elution at different positions. For example, based on its predicted structure, it appears that the 5′ TR of JEV could adopt a more compact structure com-pared to the 5′ TR of ZIKV (Figure 8.2A,B). The relatively extended conformation of 5′ TR of ZIKV will have a larger hydrodynamic radius compared to that of JEV, which will cause it to elute earlier in SEC. These results lead us to believe that both RNAs are folded correctly, whereas if the RNA were denatured, we would expect the elution peak to be virtually identical for both RNAs. The MST studies demonstrate that the JEV 5′ TR interacts with high affinity with DDX3X, compared to the ZIKV 5′ TR. As both ncRNAs are of same size but have different conformations, as suggested by SEC elution profiles (Figure 8.2C), it could be speculated that DDX3X has high affinity for compact conformation, compared to an extended one. However, the precise mechanisms that deter-mine the specificity and the biological relevance of DDX3X-ZIKV interactions require additional work.

DEAD-box helicase DDX3X has been implicated in many viral systems as a key regulator of viral replication (51,68,69). However, we lack insights into the affinity of DDX3X for any viral RNA. Moreover, whether DDX3X unwinds viral RNA is also un-clear. Therefore, we utilized MST that has emerged as one of the ideal techniques to study biomolecular interactions (77,78). By assessing the change in fluorescent migration under the influence of an infra-red laser, we are able to quantify the binding affinity between $DDX3X_{132–607}$ and 5′ TRs. To our knowledge, this work provides the first evidence of DDX3X directly interacting with in vitro transcribed, natively purified viral terminal regions, as well as the binding affinity of DDX3X with viral RNAs. Figure 8.3B and 3C also highlight how two different viral RNA TRs can have different affinities, despite being highly conserved sequences (71). The 5′ TR of JEV

has over four-times stronger affinity for $DDX3X_{132–607}$ compared to the 5′ TR of ZIKV. Furthermore, DDX3X was identified as an interacting partner for the 5′ TR of JEV in vivo previously (51). In this study, we also discovered that similar to 5′ TR of JEV, DDX3X also interacts with 5′ TR of ZIKV, albeit with lower affinity. The essential role of DDX3X in the JEV life cycle has already been established (51), but further molecular virology studies are required to investigate if DDX3X has an impact on ZIKV replication. Figure 8.4 suggests that both RNAs can be unwound by DDX3X, potentially indicating that DDX3X may still have a significant role to play, despite the affinity difference, in ZIKV replication. We performed an experiment (presented in Figure 8.4C) to simulate a knockout mutant of DDX3X, which lacks the ATP hydrolysis activity (75). We observed that the amount of fluorescent migration appears to be reduced in the absence of ATP (Figure 8.4C), which can be attributed to the fact that $DDX3X_{132–607}$ can unwind RNA but requires ATP for efficient processing of dsRNAs (79).

8.7 Conclusions

Previous studies for DDX5, a DDX3X homolog, have suggested that DDX5 interacts with RNA with high affinity in the nanomolar ranges. For example, fluorescence anisotropy experiments have suggested that DDX5 binds to short blunt-ended RNA duplexes with an affinity of ~230 nM (80). A different group tested DDX5's ability to bind short G-quadruplexes using ELISA and obtained a K_d of 22 nM (81). Our work was performed using truncated DDX3X (as all our attempts to purify the full-length DDX3X in high amounts were unsuccessful), which could impact the binding specificity and affinity for RNA compared to the full-length DDX3X. Note that a truncated construct, similar to $DDX3X_{132–607}$, was previously described as the minimally active construct (67,81). Overall, additional work aimed at investigating the role of DDX3X in ZIKV replication could provide critical information towards potential therapeutics for these deadly viruses. In conclusion, our work provides the biochemical basis of the recognition of 5′ TRs of JEV and ZIKV by DDX3X.

8.8 Materials and Methods

8.8.1 Overexpression and Purification of $DDX3X_{132–607}$

The $DDX3X_{132–607}$ cDNA construct was cloned in pET28a, followed by its transformation in Lemo21(DE3) E. coli cells. The cells were allowed to grow in Luria-broth containing kanamycin (50 mg/mL) and chloramphenicol (100 mg/mL) at 37 °C. The next day, the culture was transferred to Terrific broth containing 5% glycerol and grown at 37 °C for 5 h. The temperature was then lowered to 20 °C for 16–18 h. Cells were harvested via centrifugation and resuspended in lysis buffer (50 mM Tris, 500mM NaCl, 10mM imidazole 3mM β-mercaptoethanol, 10 mg/mL lysozyme, 0.1% Tween-20, and 5% glycerol at pH 8.0). Following 30 min of incubation on ice, the suspension was sonicated and centrifuged at 30,000× g for 45 min. The supernatant was filtered using a 0.45 µm filter for subsequent purification using chromatography methods.

Using the ÄKTA start protein purification system (Global Life Science Solutions USA LLC, Marlborough, MA, USA) equipped with the HisTrap™ High-Performance column, we purified $DDX3X_{132–607}$ via its hexa-histidine tag. Next, we used the ÄKTA pure purification system (Global Life Science Solutions USA LLC, Marlborough, MA, USA) with a Superdex® 200 10/300 GL increase column to further purify affinity-purified protein (in 50 mM Tris, 150 mM NaCl and 3% glycerol at pH 8.0). Peak fractions representing homogenous $DDX3X_{132–607}$ were collected and concentrated using Amicon® Ultra-15 Centrifugal Filter Units (30,000 kDa molecular-wight cut-off) (Millipore Canada Ltd, Etobicoke, ON, Canada). Aliquots were stored at −80 °C. SDS-PAGE (10%) was performed by taking 10 µL of SEC-purified sample and mixed it with 2 µL SDS-loading dye and heated to 95 °C for 5 min. Following heating, samples were loaded into a 1.0 cm well PAGE casting plate (Bio-Rad Laboratories, Mississauga, ON, Canada) and ran for 1 h at 200V in 1× SDS running buffer. A molecular weight ladder was run alongside the purified sample (unstained protein molecular weight marker 116 kDa to

14.4 kDa, Bio Basic Inc., Markham, ON, Canada). Subsequently, the gel was stained with Coomassie brilliant blue (Bio Basic Inc., Markham, ON, Canada) for visualization.

8.8.2 Preparation of Non-Coding RNAs

cDNA sequences were prepared under the control of T7 RNA polymerase, with two additional G nucleotides on the 5’ end with an XbaI restriction enzyme cut site (T^CTAGA) at the 3’ end. Then, 5’ TRs of JEV and Zika construct(s) were designed based on the Genebank sequence of KT957419.1 and KU509998.3, respectively. Underlined regions represent portions to which our fluorescent oligos were designed complementary to, described in a later section. Both RNA constructs used in the experiments are listed as follows:

1. JEV 5’ TR 1-156 (51.5 kDa, 159 nts)

5’GGAGAAGUUUUAUCGUGUGAACUUCUUGGCUUAGUAUCGUUGAGAAGAAUCGAGAGAUUAGU
GCAGUUUAAACAGUUUUUUAGAACGGAAGAACAACCAUGACUAAAAAACCAGGAGGGCCCGGAAA
AAACCGGGCCAUCAAUAUGCUGAAACGCGGAU3’

2. Zika 5’ TR 1-163 (52.3 kDa, 163 nts)

5’AGUUGUUACUGUUGCUGACUCAGACUGCGACAGUUCGAGUUUGAAGCGAAAGCUAGCAACAGU
AUCAACAGGUUUUAUUUGGAUUUGGAAACGAGAGUUUCUGGUCAUGAAAAACCCAAAAAAGAAAU
CCGGAGGAUUCCGGAUUGUCAAUAUGCUAAAACG3’

Each RNA was prepared through an in vitro transcription reaction using T7 RNA polymerase (purified in-house) followed by size-exclusion chromatography purification in 1× RNA buffer (10 mM Tris pH 7.5, 100 mM NaCl, and 5 mM MgCl2) using a Superdex 200 Increase GL 10/300 (Cytiva) via an ÄKTA pure FPLC (Global Life Science Solutions USA LLC, Marlborough, MA). SEC peak fractions were analyzed via urea-polyacrylamide gel electrophoresis (Urea-PAGE). Then, we mixed 10 µL of each fraction with 2 µL of denaturing RNA loading dye and loaded into a 1.0 cm well PAGE casting plate (Bio-Rad Laboratories,

Mississauga, ON, Canada). Urea-PAGE (7.2%) was run at room temperature, 300V, for 25 min in 0.5× TBE (Tris-Borate-EDTA) buffer, followed by staining with Sybr safe (Thermofisher Scientific, Saint-Laurant, QC, Canada) and visualization. Fractions containing a single band were deemed acceptable and used in subsequent experiments.

8.8.3 Fluorescent Labeling of Flaviviral RNA TRs

The 5′ TRs were incubated on ice for 30 min in 0.1 M sodium acetate (pH 5.3) along with 2 mM final concentration potassium periodate. The reaction was quenched through the addition of 10 mM final concentration ethylene glycol followed by incubation on ice for 10 min. Following incubation, we performed two ethanol precipitations and resuspended the RNA in 0.1 M NaOAc and 10 mM fluorescein-5-thiosemicarbazide (FITC) and incubated the mixture in the dark and on ice for 16 h. Following fluorescent dye incubation, the mixture was phenol extracted (1 vol phenol:1 vol mixture) 5 times until the phenol layer was consistently colourless, indicating all free dye had been re-moved from the mixture. Finally, the resulting labelled RNA was ethanol precipitated twice and resuspended in RNA buffer.

8.8.4. RNA–Protein Interaction Studies Using Microscale Thermophoresis

A 2-fold serial dilution was performed on $DDX3X_{132–607}$ whereas the highest concentration was 19 μM. Next, a constant amount of fluorescent JEV or Zika 5′ TR was mixed into each serial dilution of $DDX3X_{132–607}$, resulting in a final concentration of 20 nM of RNA. Sample mixtures were incubated at room temperature for 10 min and then placed into Nanotemper Technologies Monolith® NT.115 instrument (Nanotemper Technologies, Munich, Germany) standard capillaries and loaded into the MST. Thermophoresis was measured at room temperature (25 °C) and performed using 20% excitation power (blue filter) for both RNAs and heated using medium MST IR laser power. Fluorescent migration used to determine K_d was measured from 4.0 to 5.0 s and normalized to initial fluorescence (−1.0 to 0 s). Three independent replicates were merged and analyzed using MO.Affinity Analysis software v2.1.3

and fit to the standard K_d fit model describing a molecular interaction with a 1:1 stoichiometry according to the law of mass action. The molarity of polyU RNA could not be determined since the fragment's lengths were variable. We used a final concentration of FITC-labeled polyU of 50 µg/mL in our negative control to achieve a similar magnitude of fluorescence. K_d is estimated by fitting Equation (8.1), where F(c) is the fraction bound at a given ligand concentration c. The unbound fraction is represented by the F_{norm} signal of the target alone, and the bound fraction represents the F_{norm} signal of the complex. The K_d is the dissociation constant and c_{target} is the final concentration of the target in the assay.

$$F(c) = Unbound + (Bound - Unbound) \times \frac{c + c_{target} + K_d - \sqrt{\left(c + c_{target} + K_d\right)^2 - 4cc_{target}}}{2c_{target}} \quad (8.1)$$

8.8.5 Helicase Assay

We input our sequences into sfold (82) using standard conditions with no maxi-mum distance between paired bases and no additional constraint information. The theoretical secondary structure was used to identify a portion of each RNA molecule that was highly double-stranded. DNA oligos with complementary sequences to the double-stranded region(s) of our RNA(s) were synthesized with a 5′ conjugated FITC fluorophore (Alpha DNA). The region of each RNA molecule to which the oligos hybridize is underlined and described above. The sequences for ZIKV 5′ TR and JEV 5′ TR oligo(s) are: 5′FITC/AACTGTCGCAGTCTGAGTCAGCAACAGTAACAAC and 5′FITC/TTTCCGGGCCCTCCTGG TTTTTTAGTCATGGTTGT, respectively.

If the RNA molecule were unwound by $DDX3X_{132–607}$, it would create an opportunity for the oligos to hybridize to the 5′ TR RNA. The reaction mixture contains 20 nM of FITC-DNA oligos, 1 µM of the RNA, and 4.25 mM of ATP. $DDX3X_{132–607}$ is added to a final concentration of 10 µM. We used BSA at the same concentration as a control for DDX3X. The ATP

dependence assay involved comparing 10 μM DDX3X$_{132–607}$ with 20 nM of FITC-DNA oligos and 1 μM of the RNA, but one set of capillaries contained 4.25 mM of ATP and another set did not. An experiment consists of 3 sets of 4 capillaries for each sample, which can then be compared to detect a change in fluorescent migration be-cause of a binding event. Our data represent the normalized magnitude of fluorescent migration differences between two sets of assay conditions. Data were processed using MO.Affinity software, which assesses the signal to noise ratio between a run with and without the protein. Signal to noise is a measure of the response amplitude that is divided by the noise of the environment, as presented in Equation (8.2) (83). If the signal to noise ratio rises above 5, the assay indicates that a binding event has occurred, and the ration >12 suggests that the assay is considered as desirable (83,84).

$$\frac{S}{N} = \frac{ResponseAmplitude}{\sqrt{\frac{\sum_i (r_i - \acute{r})^2}{n-1}}} \quad (8.2)$$

8.9 References

1. Chambers, T.J. and Monath, T.P. (2003) *The flaviviruses: Detection, diagnosis and vaccine development*. Elsevier.
2. Brasil, P., Sequeira, P.C., Freitas, A.D.A., Zogbi, H.E., Calvet, G.A., De Souza, R.V., Siqueira, A.M., de Mendonca, M.C.L., Nogueira, R.M.R. and de Filippis, A.M.B. (2016) Guillain-Barré syndrome associated with Zika virus infection. *The lancet*, **387**, 1482.
3. Solomon, T., Kneen, R., Dung, N.M., Khanh, V.C., Thuy, T.T.N., Day, N.P., Nisalak, A., Vaughn, D.W. and White, N.J. (1998) Poliomyelitis-like illness due to Japanese encephalitis virus. *The Lancet*, **351**, 1094-1097.
4. Fernández-Sanlés, A., Ríos-Marco, P., Romero-López, C. and Berzal-Herranz, A. (2017) Functional information stored in the conserved structural RNA domains of flavivirus genomes. *Frontiers in microbiology*, **8**, 546.
5. Fischer, C., de Oliveira-Filho, E.F. and Drexler, J.F. (2020) Viral emergence and immune interplay in flavivirus vaccines. *The Lancet Infectious Diseases*, **20**, 15-17.
6. Organization, W.H. (2020) Epidemic Focus - Flavivirus Epidemics.
7. Pierson, T.C. and Diamond, M.S. (2020) The continued threat of emerging flaviviruses. *Nature microbiology*, **5**, 796-812.
8. Huang, Y.-J.S., Higgs, S., Horne, K.M. and Vanlandingham, D.L. (2014) Flavivirus-mosquito interactions. *Viruses*, **6**, 4703-4730.
9. Benedict, M.Q., Levine, R.S., Hawley, W.A. and Lounibos, L.P. (2007) Spread of the tiger: global risk of invasion by the mosquito Aedes albopictus. *Vector-borne and zoonotic Diseases*, **7**, 76-85.
10. Reinhold, J.M., Lazzari, C.R. and Lahondère, C. (2018) Effects of the environmental temperature on Aedes aegypti and Aedes albopictus mosquitoes: a review. *Insects*, **9**, 158.
11. Baig, S., Fox, K.K., Jee, Y., O'Connor, P., Hombach, J., Wang, S.A., Hyde, T., Fischer, M. and Hills, S.L. (2013) Japanese encephalitis surveillance and immunization—Asia and the Western Pacific, 2012. *Morbidity and Mortality Weekly Report*, **62**, 658.
12. Campbell, G.L., Hills, S.L., Fischer, M., Jacobson, J.A., Hoke, C.H., Hombach, J.M., Marfin, A.A., Solomon, T., Tsai, T.F. and Tsu, V.D. (2011) Estimated global incidence of Japanese encephalitis: a systematic review. *Bulletin of the World Health Organization*, **89**, 766-774.
13. Dubischar, K.L., Kadlecek, V., Sablan, B., Borja-Tabora, C.F., Gatchalian, S., Eder-Lingelbach, S., Kiermayr, S., Spruth, M. and Westritschnig, K. (2017) Immunogenicity of the inactivated Japanese encephalitis virus vaccine IXIARO in children from a Japanese encephalitis virus-endemic region. *The Pediatric infectious disease journal*, **36**, 898-904.
14. Van Gessel, Y., Klade, C.S., Putnak, R., Formica, A., Krasaesub, S., Spruth, M., Cena, B., Tungtaeng, A., Gettayacamin, M. and Dewasthaly, S. (2011) Correlation of protection against Japanese encephalitis virus and JE vaccine (IXIARO®) induced neutralizing antibody titers. *Vaccine*, **29**, 5925-5931.
15. Samarasekera, U. and Triunfol, M. (2016) Concern over Zika virus grips the world. *The Lancet*, **387**, 521-524.
16. Organization, W.H. (2016) Zika Situation Report.
17. Carteaux, G., Maquart, M., Bedet, A., Contou, D., Brugières, P., Fourati, S., Cleret de Langavant, L., de Broucker, T., Brun-Buisson, C. and Leparc-Goffart, I. (2016) Zika virus associated with meningoencephalitis. *New England Journal of Medicine*, **374**, 1595-1596.
18. Cauchemez, S., Besnard, M., Bompard, P., Dub, T., Guillemette-Artur, P., Eyrolle-Guignot, D., Salje, H., Van Kerkhove, M.D., Abadie, V. and Garel, C. (2016) Association between Zika virus and microcephaly in French Polynesia, 2013–15: a retrospective study. *The Lancet*, **387**, 2125-2132.

19. Gladwyn-Ng, I., Cordón-Barris, L., Alfano, C., Creppe, C., Couderc, T., Morelli, G., Thelen, N., America, M., Bessières, B. and Encha-Razavi, F. (2018) Stress-induced unfolded protein response contributes to Zika virus–associated microcephaly. *Nature neuroscience*, **21**, 63-71.
20. Mécharles, S., Herrmann, C., Poullain, P., Tran, T.-H., Deschamps, N., Mathon, G., Landais, A., Breurec, S. and Lannuzel, A. (2016) Acute myelitis due to Zika virus infection. *The Lancet*, **387**, 1481.
21. Nascimento, O.J. and Da Silva, I.R. (2017) Guillain–Barré syndrome and Zika virus outbreaks. *Current opinion in neurology*, **30**, 500-507.
22. Melo, A.O., Malinger, G., Ximenes, R., Szejnfeld, P., Sampaio, S.A. and De Filippis, A.B. (2016) Zika virus intrauterine infection causes fetal brain abnormality and microcephaly: tip of the iceberg? *Ultrasound in obstetrics & gynecology: the official journal of the International Society of Ultrasound in Obstetrics and Gynecology*, **47**, 6-7.
23. Schwartzmann, P.V., Ramalho, L.N., Neder, L., Vilar, F.C., Ayub-Ferreira, S.M., Romeiro, M.F., Takayanagui, O.M., Dos Santos, A.C., Schmidt, A. and Figueiredo, L.T. (2017), *Mayo Clinic Proceedings*. Elsevier, Vol. 92, pp. 460-466.
24. Devhare, P., Meyer, K., Steele, R., Ray, R.B. and Ray, R. (2017) Zika virus infection dysregulates human neural stem cell growth and inhibits differentiation into neuroprogenitor cells. *Cell death & disease*, **8**, e3106-e3106.
25. Broutet, N., Krauer, F., Riesen, M., Khalakdina, A., Almiron, M., Aldighieri, S., Espinal, M., Low, N. and Dye, C. (2016) Zika virus as a cause of neurologic disorders. *New England Journal of Medicine*, **374**, 1506-1509.
26. Mlakar, J., Korva, M., Tul, N., Popović, M., Poljšak-Prijatelj, M., Mraz, J., Kolenc, M., Resman Rus, K., Vesnaver Vipotnik, T. and Fabjan Vodušek, V. (2016) Zika virus associated with microcephaly. *New England Journal of Medicine*, **374**, 951-958.
27. Tang, H., Hammack, C., Ogden, S.C., Wen, Z., Qian, X., Li, Y., Yao, B., Shin, J., Zhang, F. and Lee, E.M. (2016) Zika virus infects human cortical neural progenitors and attenuates their growth. *Cell stem cell*, **18**, 587-590.
28. Li, C., Xu, D., Ye, Q., Hong, S., Jiang, Y., Liu, X., Zhang, N., Shi, L., Qin, C.-F. and Xu, Z. (2016) Zika virus disrupts neural progenitor development and leads to microcephaly in mice. *Cell stem cell*, **19**, 120-126.
29. Li, H., Saucedo-Cuevas, L., Shresta, S. and Gleeson, J.G. (2016) The neurobiology of Zika virus. *Neuron*, **92**, 949-958.
30. Russo, F.B., Jungmann, P. and Beltrão-Braga, P.C.B. (2017) Zika infection and the development of neurological defects. *Cellular microbiology*, **19**, e12744.
31. Barzon, L., Pacenti, M., Franchin, E., Lavezzo, E., Trevisan, M., Sgarabotto, D. and Palù, G. (2016) 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. *Eurosurveillance*, **21**, 30316.
32. Foy, B.D., Kobylinski, K.C., Foy, J.L.C., Blitvich, B.J., da Rosa, A.T., Haddow, A.D., Lanciotti, R.S. and Tesh, R.B. (2011) Probable non–vector-borne transmission of Zika virus, Colorado, USA. *Emerging infectious diseases*, **17**, 880.
33. Paul, L.M., Carlin, E.R., Jenkins, M.M., Tan, A.L., Barcellona, C.M., Nicholson, C.O., Michael, S.F. and Isern, S. (2016) Dengue virus antibodies enhance Zika virus infection. *Clinical & translational immunology*, **5**, e117.
34. Myers, T.R., McCarthy, N.L., Panagiotakopoulos, L. and Omer, S.B. (2019), *Open Forum Infectious Diseases*. Oxford University Press US, Vol. 6, pp. ofz071.
35. Stettler, K., Beltramello, M., Espinosa, D.A., Graham, V., Cassotta, A., Bianchi, S., Vanzetta, F., Minola, A., Jaconi, S. and Mele, F. (2016) Specificity, cross-reactivity, and function of antibodies elicited by Zika virus infection. *Science*, **353**, 823-826.

36. Cleaves, G.R. and Dubin, D.T. (1979) Methylation status of intracellular dengue type 2 40 S RNA. *Virology*, **96**, 159-165.
37. Brinton, M.A., Fernandez, A.V. and Dispoto, J.H. (1986) The 3'-nucleotides of flavivirus genomic RNA form a conserved secondary structure. *Virology*, **153**, 113-121.
38. Wengler, G., Wengler, G. and Gross, H.J. (1978) Studies on virus-specific nucleic acids synthesized in vertebrate and mosquito cells infected with flaviviruses. *Virology*, **89**, 423-437.
39. Gritsun, T. and Gould, E. (2006) Origin and evolution of 3' UTR of flaviviruses: long direct repeats as a basis for the formation of secondary structures and their significance for virus transmission. *Advances in virus research*, **69**, 203-248.
40. Thurner, C., Witwer, C., Hofacker, I.L. and Stadler, P.F. (2004) Conserved RNA secondary structures in Flaviviridae genomes. *Journal of General Virology*, **85**, 1113-1124.
41. Cross, R., Jackson, A., Citi, S., Kendrick-Jones, J. and Bagshaw, C. (1988) Active site trapping of nucleotide by smooth and non-muscle myosins. *Journal of molecular biology*, **203**, 173-181.
42. Hahn, C.S., Hahn, Y.S., Rice, C.M., Lee, E., Dalgarno, L., Strauss, E.G. and Strauss, J.H. (1987) Conserved elements in the 3' untranslated region of flavivirus RNAs and potential cyclization sequences. *Journal of molecular biology*, **198**, 33-41.
43. Alvarez, D.E., Lodeiro, M.F., Luduena, S.J., Pietrasanta, L.I. and Gamarnik, A.V. (2005) Long-range RNA-RNA interactions circularize the dengue virus genome. *Journal of virology*, **79**, 6631-6643.
44. Li, X.-D., Deng, C.-L., Yuan, Z.-M., Ye, H.-Q. and Zhang, B. (2020) Different degrees of 5'-to-3'DAR interactions modulate Zika virus genome cyclization and host-specific replication. *Journal of virology*, **94**, e01602-01619.
45. de Borba, L., Villordo, S.M., Iglesias, N.G., Filomatori, C.V., Gebhard, L.G. and Gamarnik, A.V. (2015) Overlapping local and long-range RNA-RNA interactions modulate dengue virus genome cyclization and replication. *Journal of virology*, **89**, 3430-3437.
46. Khromykh, A.A., Meka, H., Guyatt, K.J. and Westaway, E.G. (2001) Essential role of cyclization sequences in flavivirus RNA replication. *Journal of virology*, **75**, 6719-6728.
47. Friebe, P., Shi, P.-Y. and Harris, E. (2011) The 5' and 3' downstream AUG region elements are required for mosquito-borne flavivirus RNA replication. *Journal of virology*, **85**, 1900-1905.
48. Chambers, T.J., Hahn, C.S., Galler, R. and Rice, C.M. (1990) Flavivirus genome organization, expression, and replication. *Annual review of microbiology*, **44**, 649-688.
49. Tingting, P., Caiyun, F., Zhigang, Y., Pengyuan, Y. and Zhenghong, Y. (2006) Subproteomic analysis of the cellular proteins associated with the 3' untranslated region of the hepatitis C virus genome in human liver cells. *Biochemical and biophysical research communications*, **347**, 683-691.
50. Harris, D., Zhang, Z., Chaubey, B. and Pandey, V.N. (2006) Identification of cellular factors associated with the 3'-nontranslated region of the hepatitis C virus genome. *Molecular & Cellular Proteomics*, **5**, 1006-1018.
51. Li, C., Ge, L.-l., Li, P.-p., Wang, Y., Dai, J.-j., Sun, M.-x., Huang, L., Shen, Z.-q., Hu, X.-c. and Ishag, H. (2014) Cellular DDX3 regulates Japanese encephalitis virus replication by interacting with viral un-translated regions. *Virology*, **449**, 70-81.
52. Li, C., Ge, L.-l., Li, P.-p., Wang, Y., Sun, M.-x., Huang, L., Ishag, H., Di, D.-d., Shen, Z.-q. and Fan, W.-x. (2013) The DEAD-box RNA helicase DDX5 acts as a positive regulator of Japanese encephalitis virus replication by binding to viral 3' UTR. *Antiviral research*, **100**, 487-499.
53. Ward, A.M., Bidet, K., Yinglin, A., Ler, S.G., Hogue, K., Blackstock, W., Gunaratne, J. and Garcia-Blanco, M.A. (2011) 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 biology*, **8**, 1173-1186.

54. Dong, Y., Yang, J., Ye, W., Wang, Y., Miao, Y., Ding, T., Xiang, C., Lei, Y. and Xu, Z. (2015) LSm1 binds to the Dengue virus RNA 3'UTR and is a positive regulator of Dengue virus replication. *International Journal of Molecular Medicine*, **35**, 1683-1689.
55. Gomila, R.C., Martin III, G.W. and Gehrke, L. (2011) NF90 binds the dengue virus RNA 3' terminus and is a positive regulator of dengue virus replication. *PloS one*, **6**, e16687.
56. Ward, A.M., Calvert, M.E., Read, L.R., Kang, S., Levitt, B.E., Dimopoulos, G., Bradrick, S.S., Gunaratne, J. and Garcia-Blanco, M.A. (2016) The Golgi associated ERI3 is a Flavivirus host factor. *Scientific reports*, **6**, 34379.
57. Lei, Y., Huang, Y., Zhang, H., Yu, L., Zhang, M. and Dayton, A. (2011) Functional interaction between cellular p100 and the dengue virus 3' UTR. *Journal of General Virology*, **92**, 796-806.
58. Yu, L., Nomaguchi, M., Padmanabhan, R. and Markoff, L. (2008) Specific requirements for elements of the 5' and 3' terminal regions in flavivirus RNA synthesis and viral replication. *Virology*, **374**, 170-185.
59. Meier-Stephenson, V., Mrozowich, T., Pham, M. and Patel, T.R. (2018) DEAD-box helicases: the Yin and Yang roles in viral infections. *Biotechnology and Genetic Engineering Reviews*, **34**, 3-32.
60. Ng, W.C., Soto-Acosta, R., Bradrick, S.S., Garcia-Blanco, M.A. and Ooi, E.E. (2017) The 5' and 3' untranslated regions of the flaviviral genome. *Viruses*, **9**, 137.
61. Leitão, A.L., Costa, M.C. and Enguita, F.J. (2015) Unzippers, resolvers and sensors: a structural and functional biochemistry tale of RNA helicases. *Int J Mol Sci*, **16**, 2269-2293.
62. Linder, P. and Jankowsky, E. (2011) From unwinding to clamping—the DEAD box RNA helicase family. *Nature reviews Molecular cell biology*, **12**, 505-516.
63. Pyle, A.M. (2008) Translocation and unwinding mechanisms of RNA and DNA helicases. *Annu. Rev. Biophys.*, **37**, 317-336.
64. Jankowsky, E. (2011) RNA helicases at work: binding and rearranging. *Trends in biochemical sciences*, **36**, 19-29.
65. Chao, C.-H., Chen, C.-M., Cheng, P.-L., Shih, J.-W., Tsou, A.-P. and Wu Lee, Y.-H. (2006) DDX3, a DEAD box RNA helicase with tumor growth–suppressive property and transcriptional regulation activity of the p21waf1/cip1 promoter, is a candidate tumor suppressor. *Cancer research*, **66**, 6579-6588.
66. Ariumi, Y., Kuroki, M., Abe, K.-i., Dansako, H., Ikeda, M., Wakita, T. and Kato, N. (2007) DDX3 DEAD-box RNA helicase is required for hepatitis C virus RNA replication. *Journal of virology*, **81**, 13922-13926.
67. Song, H. and Ji, X. (2019) The mechanism of RNA duplex recognition and unwinding by DEAD-box helicase DDX3X. *Nature communications*, **10**, 3085.
68. Li, G., Feng, T., Pan, W., Shi, X. and Dai, J. (2015) DEAD-box RNA helicase DDX3X inhibits DENV replication via regulating type one interferon pathway. *Biochemical and biophysical research communications*, **456**, 327-332.
69. Brai, A., Martelli, F., Riva, V., Garbelli, A., Fazi, R., Zamperini, C., Pollutri, A., Falsitta, L., Ronzini, S. and Maccari, L. (2019) DDX3X helicase inhibitors as a new strategy to fight the West Nile virus infection. *Journal of medicinal chemistry*, **62**, 2333-2347.
70. Brinton, M.A. (2014) Replication cycle and molecular biology of the West Nile virus. *Viruses*, **6**, 13-53.
71. Brinton, M.A. and Basu, M. (2015) Functions of the 3' and 5' genome RNA regions of members of the genus Flavivirus. *Virus research*, **206**, 108-119.
72. Reuten, R., Patel, T.R., McDougall, M., Rama, N., Nikodemus, D., Gibert, B., Delcros, J.-G., Prein, C., Meier, M. and Metzger, S. (2016) Structural decoding of netrin-4 reveals a regulatory function towards mature basement membranes. *Nature communications*, **7**, 13515.

73. Ferens, F.G., Patel, T.R., Oriss, G. and Stetefeld, J. (2019) A cholesterol analog induces an oligomeric reorganization of VDAC. *Biophysical journal*, **116**, 847-859.
74. Wienken, C.J., Baaske, P., Rothbauer, U., Braun, D. and Duhr, S. (2010) Protein-binding assays in biological liquids using microscale thermophoresis. *Nature communications*, **1**, 100.
75. Garbelli, A., Beermann, S., Di Cicco, G., Dietrich, U. and Maga, G. (2011) A motif unique to the human DEAD-box protein DDX3 is important for nucleic acid binding, ATP hydrolysis, RNA/DNA unwinding and HIV-1 replication. *PloS one*, **6**, e19810.
76. Nelson, C.R., Mrozowich, T., Park, S.M., D'souza, S., Henrickson, A., Vigar, J.R., Wieden, H.-J., Owens, R.J., Demeler, B. and Patel, T.R. (2020) Human DDX17 unwinds rift valley fever virus non-coding RNAs. *Int J Mol Sci*, **22**, 54.
77. Jerabek-Willemsen, M., Wienken, C.J., Braun, D., Baaske, P. and Duhr, S. (2011) Molecular interaction studies using microscale thermophoresis. *Assay and drug development technologies*, **9**, 342-353.
78. Mrozowich, T., MeierStephenson, V. and Patel, T.R. (2019) Microscale thermophoresis: warming up to a new biomolecular interaction technique. *The Biochemist*, **41**, 8-12.
79. Liu, F., Putnam, A. and Jankowsky, E. (2008) ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. *Proceedings of the National Academy of Sciences*, **105**, 20209-20214.
80. Xing, Z., Wang, S. and Tran, E.J. (2017) Characterization of the mammalian DEAD-box protein DDX5 reveals functional conservation with S. cerevisiae ortholog Dbp2 in transcriptional control and glucose metabolism. *RNA*, **23**, 1125-1138.
81. Floor, S.N., Condon, K.J., Sharma, D., Jankowsky, E. and Doudna, J.A. (2016) Autoinhibitory interdomain interactions and subfamily-specific extensions redefine the catalytic core of the human DEAD-box protein DDX3. *Journal of Biological Chemistry*, **291**, 2412-2421.
82. Ding, Y., Chan, C.Y. and Lawrence, C.E. (2004) S fold web server for statistical folding and rational design of nucleic acids. *Nucleic acids research*, **32**, W135-W141.
83. Technologies, N. (2018) User Manual for the Monolith NT.115.
84. Seidel, S.A., Dijkman, P.M., Lea, W.A., van den Bogaart, G., Jerabek-Willemsen, M., Lazic, A., Joseph, J.S., Srinivasan, P., Baaske, P. and Simeonov, A. (2013) Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. *Methods*, **59**, 301-315.

Chapter 9. Investigating RNA-RNA interactions through computational and biophysical analysis

9.1 Foreword

This chapter consists of a manuscript I wrote in collaboration Sean Park, whom I share co-authorship with, Maria Waldl, Amy Henrickson, Scott Tersteeg, Corey Nelson, Anneke Deklerk, Dr. Borries Demeler, Dr. Ivo Hofacker, Dr. Michael Wolfinger and Dr. Trushar R Patel. As of the thesis submission date, this manuscript has undergone a first round of revisions in *Nucleic Acids Research* and had been resubmitted. It is additionally online in preprint at https://doi.org/10.1101/2022.02.01.478553. This work focuses on the interaction of JEV 5' and 3' non-coding terminal regions using a combination of biophysical and computational methods. Sean Park assisted with all experiments. Maria Waldl contributed the computational methods, Corey Nelson assisted with MST data analysis while Scott Tersteeg, and Anneke Deklerk helped produce mutant RNA for first-round revisions. Amy Henrickson assisted with analytical ultracentrifugation. Sean additionally assisted with writing of the final draft. This publication is reproduced with permission from Sean Park and the Creative Commons agreement (open access) and is re-formatted to fit thesis formatting.

9.2 Author list and Affiliations

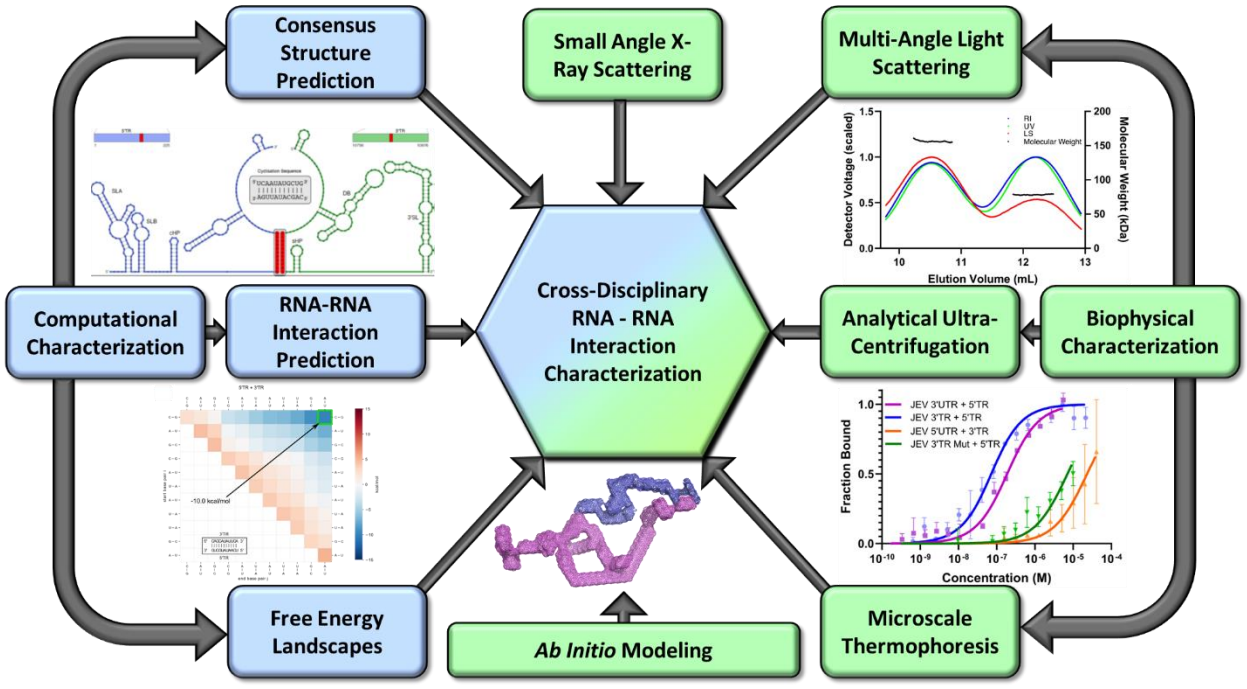
Tyler Mrozowich[1†], Sean M. Park[1†,] Maria Waldl[2,3], Amy Henrickson[1], Scott Tersteeg[1], Corey R. Nelson[1], Anneke Deklerk[1], Borries Demeler[1,4,5], Ivo L. Hofacker[2,6], Michael T. Wolfinger[2,6], Trushar R. Patel[1,7,8]

1. Department of Chemistry and Biochemistry, Alberta RNA Research and Training Institute, University of Lethbridge, 4401 University Drive, Lethbridge, AB T1K 3M4, Canada
2. Department of Theoretical Chemistry, University of Vienna, Währinger Strasse 17, 1090, Vienna, Austria
3. Center of Anatomy & Cell Biology, Division of Cell & Developmental Biology, Medical, University of Vienna, Schwarzspanierstrasse 17, 1090, Vienna, Austria
4. Department of Chemistry and Biochemistry, University of Montana, Missoula, MT 59812, USA
5. NorthWest Biophysics Consortium, University of Lethbridge, University of Lethbridge, 4401 University Drive, Lethbridge, AB T1K 3M4, Canada
6. Bioinformatics and Computational Biology, Faculty of Computer Science, University of Vienna, Währinger Strasse 29, 1090, Vienna Austria
7. Li Ka Shing Institute of Virology, University of Alberta, Edmonton, T6G 2E1, Alberta, Canada
8. Department of Microbiology, Immunology & Infectious Diseases, Cumming School of Medicine, University of Calgary, Calgary, T2N 4N1, Canada

† Equal author contribution

9.3 Abstract

Numerous viruses utilize essential long-range RNA-RNA genome interactions, specifically flaviviruses. Using Japanese encephalitis virus as a model system, we computationally predicted and then biophysically validated and characterized its long-range RNA-RNA genomic interaction. Using multiple RNA computation assessment programs, we determine the primary RNA-RNA interacting site among JEV isolates and numerous related viruses. Following *in vitro* transcription of RNA, we provide, for the first time, characterization of an RNA-RNA interaction using multi-angle light scattering and analytical ultra-centrifugation. Next, we demonstrate that the 5' and 3' TRs of JEV interact with nM affinity using microscale thermophoresis (MST), and this affinity is significantly reduced when the conserved cyclization sequence is not present. Furthermore, we perform computational kinetic analyses validating the cyclization sequence as the primary driver of this RNA-RNA interaction.



Graphical Abstract

Finally, we examined the 3-dimensional structure of the interaction using small-angle X-ray

scattering, revealing a flexible yet stable interaction. This pathway can be adapted and utilized to study various viral and human long-non-coding RNA-RNA interactions and determine their binding affinities, a critical pharmacological property of designing potential therapeutics.

9.4 Introduction

Computational and biophysical methods to obtain structure, properties, and dynamics of protein-ligand interactions are well developed as a foundational component of many pharmaceutical discovery pipelines. As a result, nearly all currently approved drugs target one of ~700 disease-related proteins, despite an increasing number of diseases attributed to the >98% of the human genome, which is non-coding (1). Amongst the limitations in targeting these elements is a lack of foundational techniques for characterizing RNA-RNA interactions in the same depth as protein-ligand interactions. The importance of RNA characterization has only gained traction as of late, highlighting an ever-increasing need for techniques capable of doing so (2-5). In this study, we describe the novel application of well-established biophysical modalities with computational studies for extensive characterization of RNA-RNA interactions using a flaviviral system theorized to contain crucial intragenomic RNA-RNA interactions, Japanese encephalitis virus.

JEV is a mosquito-borne flavivirus in the genus Flavivirus (family Flaviviridae), which contains several pathogenic viruses such as Dengue virus, West Nile virus, Zika virus, yellow fever virus. JEV is the leading cause of viral encephalitis in Southeast Asia and the Western Pacific, with approximately 68,000 cases globally each year (6). Currently, no approved treatments are available following infection by JEV or other flaviviral infections (7), which lends to the importance of further investigation. JEV is transmitted through competent mosquito vectors of the genera Aedes and Culex (8), suggesting that flaviviral infections will likely become more prevalent as global temperatures rise and vector populations expand (9). Like other flaviviruses, JEV is an enveloped virus with a single-stranded (+)-sense RNA genome of approximately 11,000nt in length (10). A single open reading frame encodes for a

polyprotein and is flanked by highly structured 5′ and 3′ untranslated regions (11). The genome has a type I cap at the 5′ end (m7GpppAMP) and lacks polyadenylation at the 3′-terminus. The single open reading frame is cleaved post-translationally into three structural and seven non-structural proteins (12,13). During replication, the 5′ and 3′ terminal regions in flaviviruses undergo long-range intragenomic RNA-RNA interactions, thereby forming a so-called panhandle structure that mediates recruitment of the viral RNA-dependent RNA polymerase (NS5) (14). Removal of the TRs has shown inhibition of viral replication (14-17). A cyclization sequence of 11nt is complementary in the 5′ and 3′ TRs, which facilitates this interaction (18). Furthermore, these terminal regions show binding with a variety of human host proteins (19), including but not limited to numerous DEAD-box helicases (20). In WNV and DENV, the 5′-3′ long-range interaction has been previously demonstrated (16-18,21,22). Genome cyclization in JEV has been computationally predicted previously (23); however, detailed experimental verification is still missing.

In this study, we explore the novel application of complementary computational and biophysical techniques that have not been used to characterize RNA-RNA interactions. Through a computational approach, we first identified an isolate of JEV that we hypothesized to interact with high affinity, then evaluated a consensus duplex structure of the 5′ and 3′ TRs of 20 different flaviviruses and measured their conservation. With this knowledge and utilizing various biophysical characterization techniques, we directly demonstrate for the first time that JEV 5′-3′ TRs interact *in vitro* with nanomolar affinity and with 1:1 stoichiometry. Furthermore, we isolated and identified the RNA-RNA complex using size exclusion chromatography coupled multi-angle light scattering. We additionally use analytical ultra-centrifugation as an orthogonal biophysical validation as evidence of JEV 5′-3′TR interaction. We provide computational evidence which complements our experimental data showing that the cyclization sequence interaction is kinetically favorable and will out-compete potential homo-dimer RNA interactions. Finally, we present a low-resolution *ab inito* model showing the

potential architectural arrangement of this RNA-RNA interaction in solution. This characterization can be used as a foundation in potential pharmaceutical therapies to inhibit viral replication through cyclization interruption and to help understand the replication pathway/mechanism by which NS5 replicates the viral genome.

9.5 Results and Discussion

9.5.1 Computational analysis of the cyclization RNA-RNA interacting element in JEV and related flaviviruses

We performed computational analysis of the potential long-range interaction between the 5' and 3’ TRs in more than 100 JEV isolates. This work revealed that all isolates interact via the canonical 11nt cyclization sequences in their 5’ and 3’ TRs. Based on this work, we selected JEV isolate KR265316.1 as a model to assess long-range RNA-RNA interactions *in*

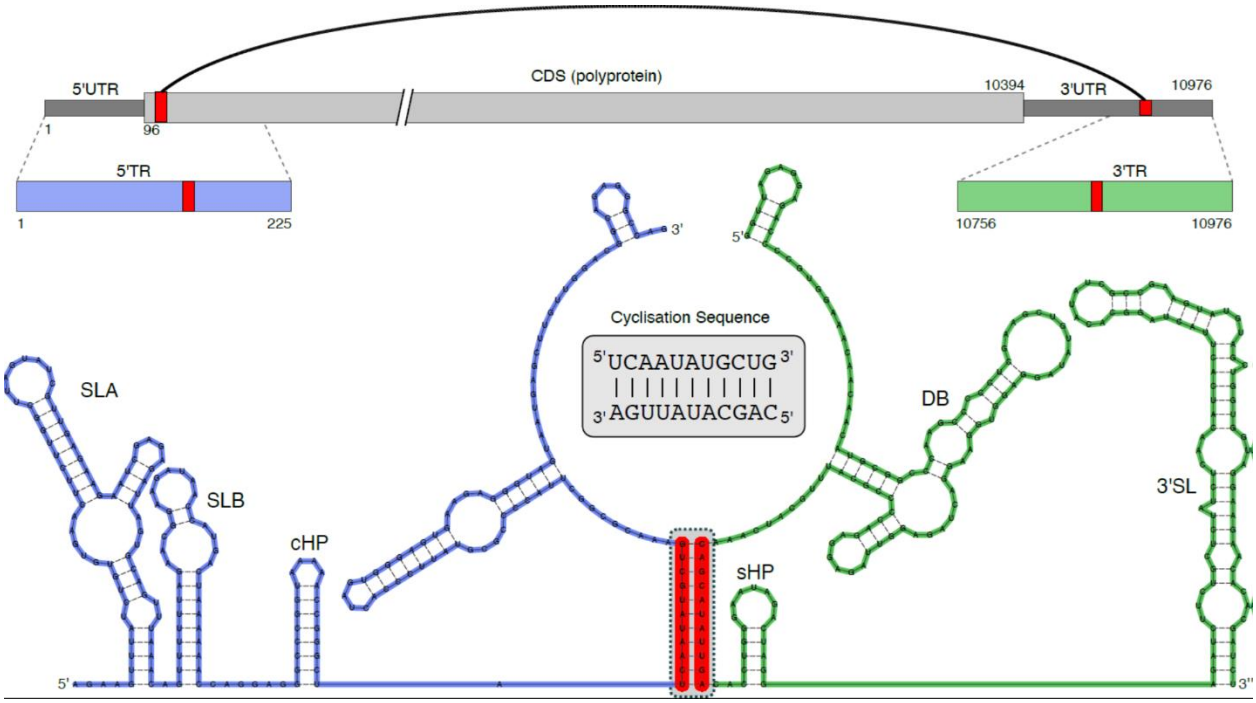


Figure 9.1. JEV structural organization and theoretical secondary structure prediction. Visual representation of JEV genome organization and different construct regions which were designed and studied. Predicted secondary structure of the cyclization between JEV 5’TR and JEV 3’TR from constraining co-folding, highlighted red portions represents the 11 conserved nucleotides which form a stable base pairing interaction to facilitate the long-range genome interaction.

silico and *in vitro*. The Flaviviral TRs are known to harbor functional RNA elements such as

stem-loops A and B in the 5′UTR, a short conserved hairpin (cHP) at the beginning of the coding region, and multiple cis-regulatory elements in the 3′UTR, such as dumbbell and terminal 3′ stem-loop (3′SL) structures (24,25). These conserved elements exert crucial roles in the flaviviral life cycle; therefore, we required them to be formed by constraint co-folding of the 5′ and 3 ′TRs. As presented in Figure 9.1, the resulting duplex structure suggests that the canonical 11nt cyclization sequence in the 5′ and 3′ TRs is responsible for mediating the interaction. Furthermore, the other canonical secondary structures, such as SLA, SLB, and cHP in the 5′ TR with sHP, the 3′SL, and the 3′DB in the 3′ TR, are also present (Figure 9.1).

We performed a comparative genomics assay in phylogenetically related viruses to further assess flaviviral long-range RNA-RNA interactions. To this end, we analyzed the

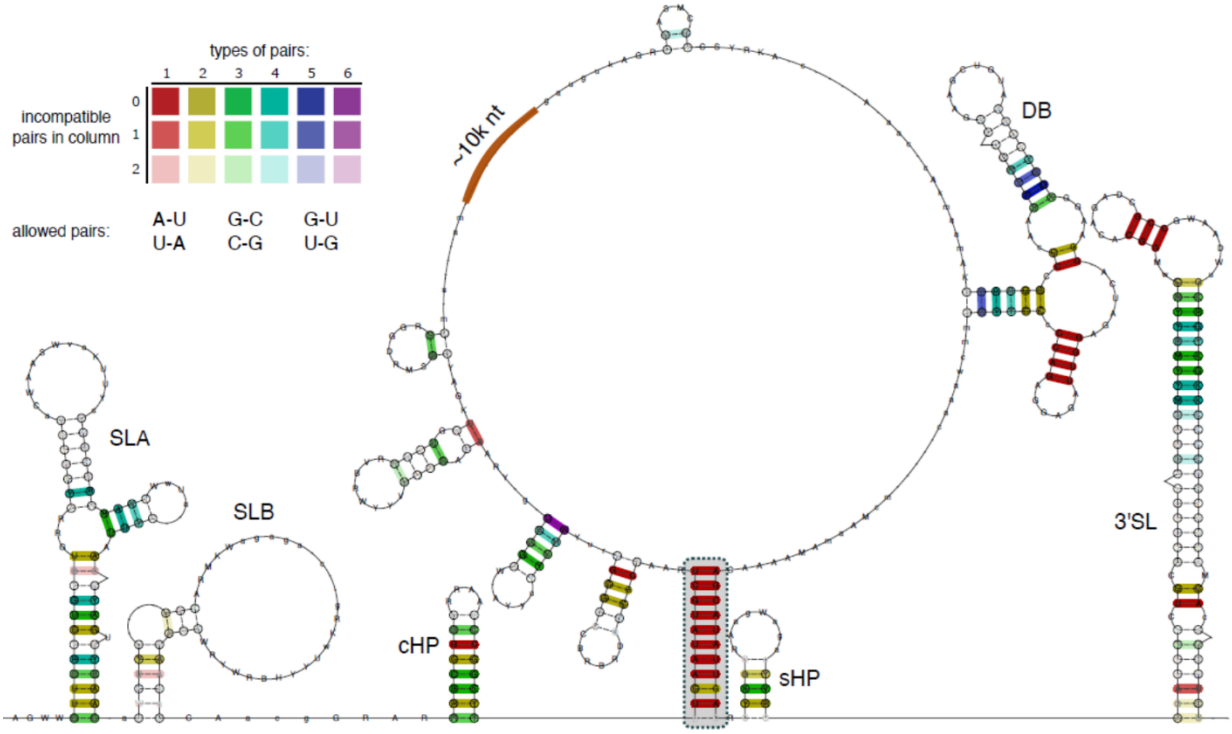


Figure 9.2. Consensus secondary structure of the 5'TR/3'TR long-range interaction. Computed from a structural multiple sequence alignment of 20 mosquito-borne flaviviruses. (Figure S9.1). Coloring of base pairs follows the RNAalifold schema and indicates different covariation levels, ranging from red (no covariation, full primary sequence conservation) to violet (full covariation, all six possible combinations of base pairs) at corresponding columns of the underlying alignment. The duplex formed by the almost fully sequence-conserved 11 nt cyclization sequences (highlighted in gray) represents the only long-range interaction in the consensus structure. Canonical 5'TR elements (SLA, SLB, and cHP), as well as 3'TR elements (DB, sHP and 3'SL) are predicted to fold in the consensus structure, indicating that they are energetically more favourable than an extended long-range interaction duplex structure.

propensity of duplex formation between 5′ and 3′TRs in 20 mosquito-borne flaviviruses (Figure

S9.2) utilizing consensus structure evaluation of the terminal genomic regions (Figure 9.2). Interestingly, the tendency to form a long-range interaction via the canonical cyclization motif is more pronounced in the consensus duplex than in the single sequence (JEV only) analysis. Specifically, the 11bp duplex is formed in the consensus structure without any constraints on the known 5' and 3' UTR elements (Figure 9.2). All 20 sampled mosquito-borne flaviviruses have the capacity to form the exact same long-range interaction. The almost perfect conservation of the cyclization sequences suggests that this region is not only functionally essential but also suggests that there may be evolutionary importance to the sequence conservation.

Additionally, a kinetic analysis of the canonical cyclization structure suggests that it is also kinetically favoured in all investigated JEV isolates. Exemplarily, the energy landscape of the cyclization structure from isolate KR265316.1 is shown in Figure 9.6 (discussed later). Independent of the selected start base pair, every interaction extension step leads to a more stable structure. Thus, there are no barriers along the folding paths, and the interaction can form fast. We then sought to validate the *in silico*-derived cyclization interaction in JEV

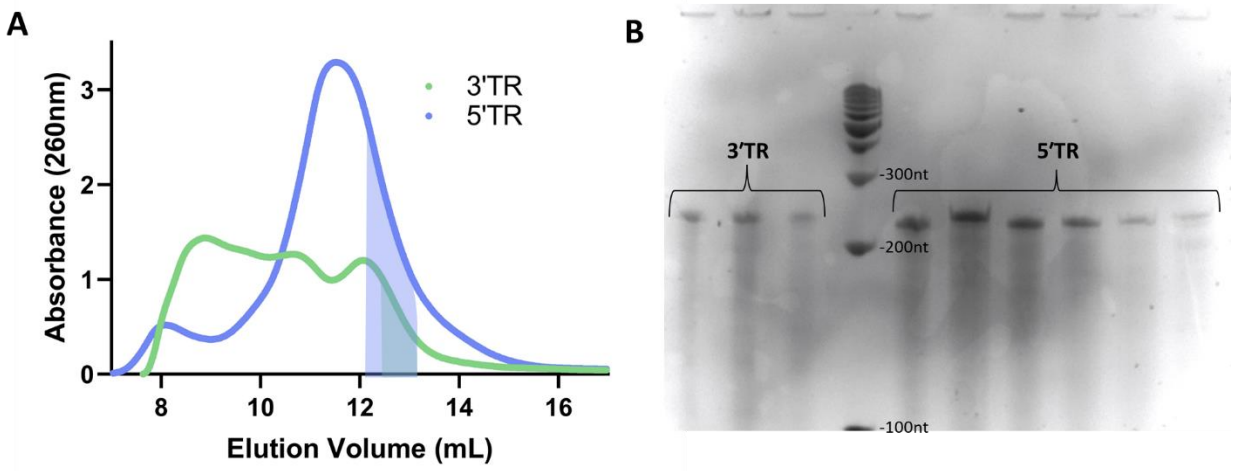


Figure 9.3. Purification of JEV TR RNA A) Size Exclusion Chromatogram representing the purification of both RNA. Shaded boxes represent the region which was collected for downstream experiments to avoid any potential oligomeric species. **B)** Urea PAGE of associated size exclusion chromatography fractions showing a single size of RNA (~220nt) which is the correct size of the expected RNA.

through extensive biophysical characterization *in vitro*.

9.5.2 In vitro transcription and purification of JEV non-coding RNA for interaction studies

JEV viral non-coding RNAs were purified immediately after *in vitro* transcription using SEC, like in previous works (20,26,27). The elution profile for 5'TR (Figure 9.3A) indicates that the RNA elutes at approximately 11.5 mL as a single monodispersed species, evident by the typical Gaussian distribution. 3'TR elutes as a multimodal distribution with an observable peak at ~12 mL, consistent with its size compared to 5'TR (Figure 9.1A). The 5'TR and 3'TR transcripts have very similar molecular weights, 73.5 kDa and 71.3 kDa, respectively, which is reflected in the chromatogram as both monomeric peaks appear to have a similar elution volume ~11.5-12.0 mL (Figure 9.3A). Higher-order oligomeric species for 3'TR can be observed at ~8 -11 mL, and these elution fractions we avoided for downstream experiments.

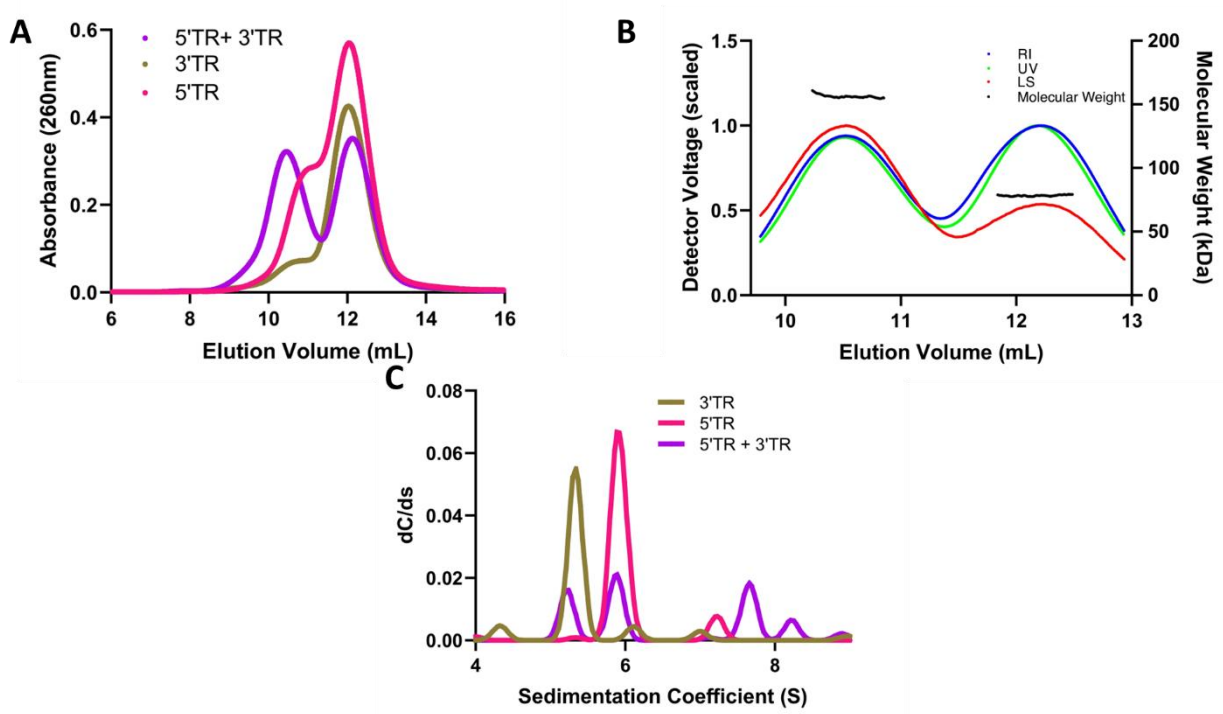


Figure 9.4. Light scattering analysis of JEV TR RNA cyclization A) Multiple size exclusion chromatography runs associated with SEC-MALS. **B)** MALS traces of each peak from the 5'TR +3'TR run, and the absolute molecular weight across them. **C)** Sedimentation distribution profiles of JEV 5' TR and 3'TR obtained from sedimentation velocity-analytical ultracentrifugation. Sedimentation coefficient values are corrected to standard solvent conditions (20°C, water)

To determine if the SEC elution fractions contained the correct-sized RNA species, we utilized urea-PAGE. Figure 9.3B (left side) shows 3'TR fractions from the right side of the ~12 mL

peak containing the appropriate length RNA species (227 nt). Only selected fractions (12.5 – 13.5 mL) were pooled to avoid contamination from any oligomeric species (Figure 9.3A, red). 5'TR elution fractions (Figure 9.3B, right side) demonstrate, as expected, a single species of the correct RNA size (221 nt) across the peak. These fractions were pooled similarly to 3'TR and highlighted for clarification (Figure 9.3A, green). 3'TR Mut, 3'UTR, and 5'UTR (additional RNA used later) were transcribed and purified identically to the above RNA, and only fractions containing a single band were pooled and used in downstream experiments.

9.5.3 Biophysical analysis of the RNA-RNA interacting Complex.

Due to the highly conserved nature of Flavivirus TRs, it is theorized that JEV will utilize similar 3' – 5' long-range TR interactions as a necessary step in viral RNA amplification, like other members (49). This interaction has been shown in various Flaviviridae family members (16-18,21,22). SEC can separate biomolecules based on their sizes, and MALS allows for the absolute molecular weight determination of biomacromolecules in solution (28); we utilized

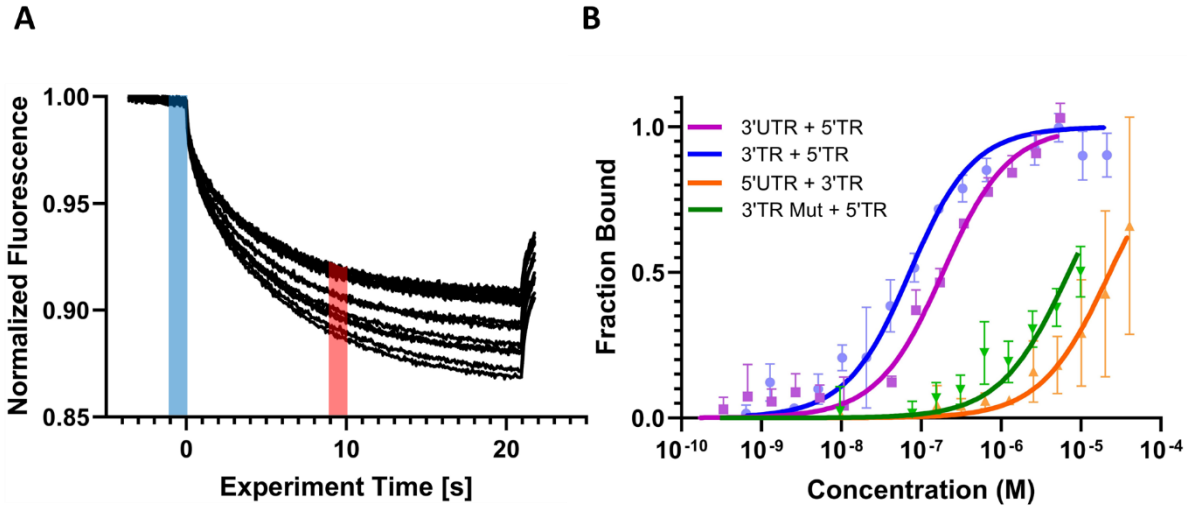


Figure 9.5. Affinity analysis of JEV TR RNA cyclization A) MST raw data traces for JEV 3'TR + 5'TR. Blue line represents the "cold" time and the red line represent the "hot" region, and the difference between the two is used to calculate the ΔF_{norm} **B)** Microscale thermophoresis measurements of different combinations of JEV TRs representing concentration vs fraction bound. Measurement ran on "high" MST power.

SEC-MALS to investigate if the 5' and 3'TRs form a complex. Figure 9.4A represents the elution profiles of JEV 3'TR, which indicates that it is almost entirely monodispersed eluting at ~12

mL with a minor peak of potential oligomeric assembly at ~11.5 mL. The elution profile of 3’ TR is consistent with our initial purification using SEC, where JEV 3’TR oligomerizes into higher-order species. JEV 5’TR also elutes at ~12 mL, consistent with previous purifications; however, it also contains a shoulder at ~11.5 mL (Figure 9.4A).

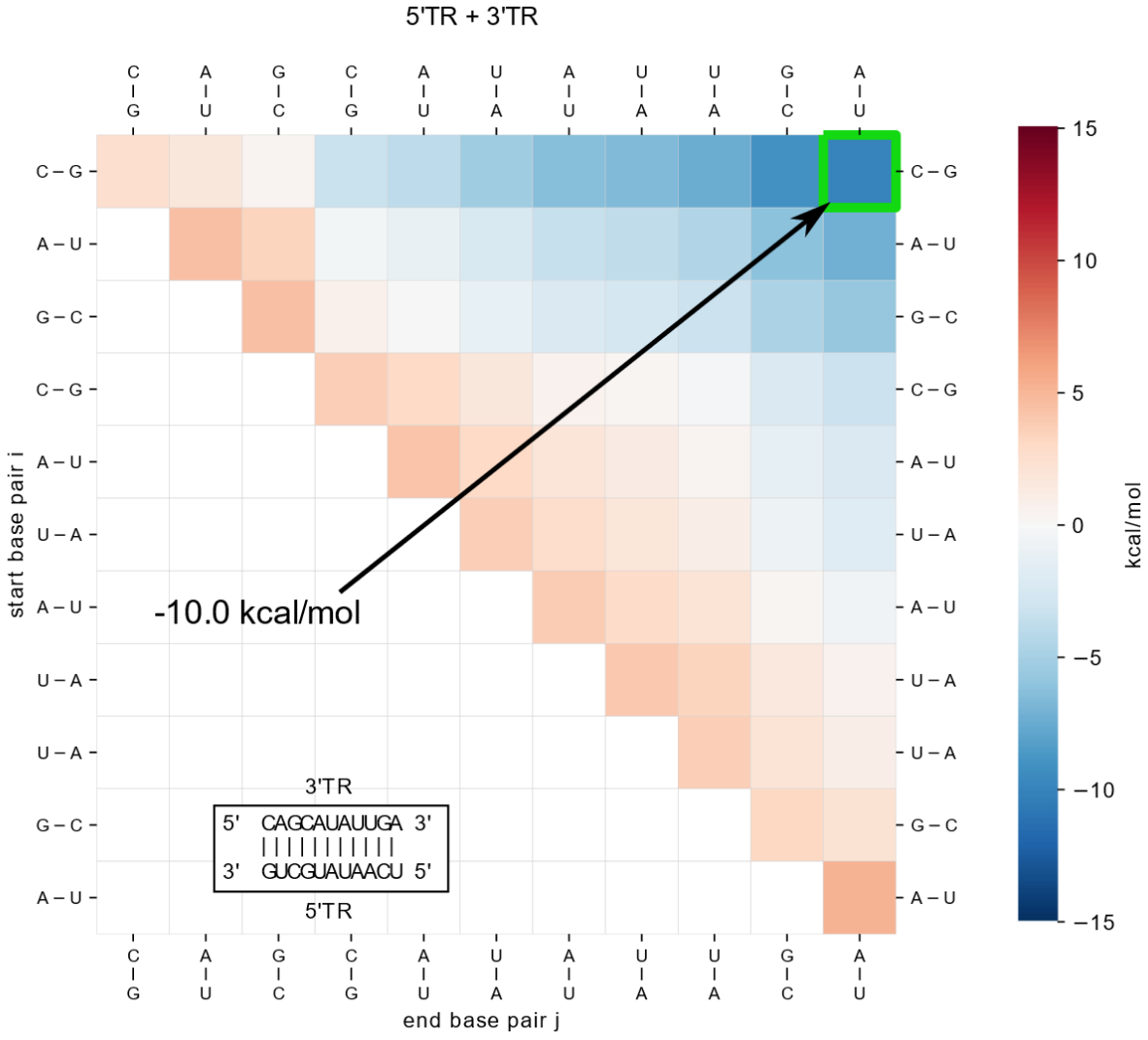


Figure 9.6. Energy landscape of the predicted 3’TR and 5’TR CS interaction. The energy landscape represents all possible substructures that can occur along a direct path from a single base pair interaction to the full CS interaction.

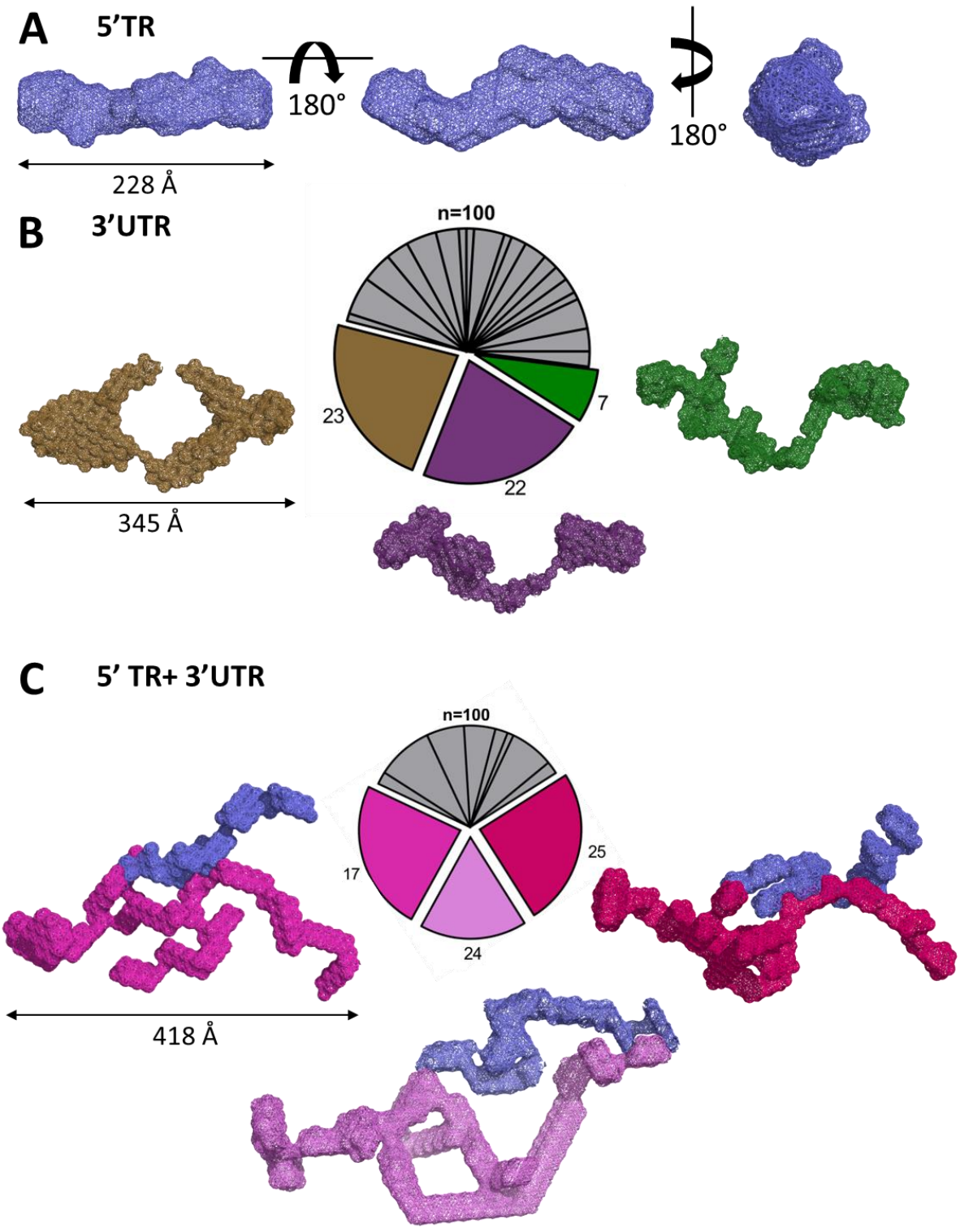


Figure 9.7. Small-angle X-ray scattering *ab inito* model reconstructions of JEV RNA. A) JEV 5' TR RNA, rotations are 180° about both the X and Y axis and showing D_{max}. **B)** JEV 3' UTR RNA illustrating relative conformational fluidity through clustering analysis of 100 *ab inito* models. Shown models represent three of the largest clusters of models. **C)** JEV 5'TR+3'UTR RNA-RNA complex. JEV 5'TR is represented in blue, while 3'UTR is represented in shades of red/purple.

We then mixed both 3'TR and 5'TR and performed SEC-MALS experiments. The RNA-RNA mix resulted in a bimodal SEC chromatogram distribution, with an additional peak eluting at ~10.5 mL (Figure 9.4A, purple). Furthermore, we see a peak at ~12 mL where excess 5'TR RNA elutes as a monomer, consistent with the concentrations we mixed (3.5 μM of JEV 5'TR vs. 1.3 μM of JEV 3'TR). These results indicate that the TR complex has a different elution profile upon binding than any oligomeric assembly or monomeric species. Successful separation of the TR RNA complex allowed us to calculate the absolute molecular weight of each species. As presented in Figure 9.4B, the SEC-MALS-derived molecular weights are consistent across both peaks, indicating that both are homogenous species. Excess unbound monomer(s) elute as a single peak and are characterized at ~75 kDa, which is expected considering the similarity in their theoretical weights (73.9 kDa & 71.4 kDa). The molecular weight of the complex (~150kDa) is double that of each TR in isolation (~75kDa) and very similar to the predicted molecular weight of the complex (145.3 kDa). This change in molecular weight is indicative of a 1:1 stoichiometric interaction. This stoichiometric determination is vital for biological relevance, considering these terminal regions exist on the same genome and come together to cyclize.

AUC is a powerful biophysical technique often used to study the purity of biomolecules in solution (29). AUC subjects biomolecules to extremely high centrifugal force (up to 250,000 x g), separating them based on size, anisotropy, and density while monitoring sedimentation via an optical system. While SEC-MALS can provide us with absolute molecular weight determination, since both 5'TR and 3'TR are of similar molecular weight, we needed further validation to rule out potential RNA self-oligomerization as an explanation of the 5'TR-3'TR peak in Figure 9.3A. AUC has been proven to be a useful technique to characterize RNA (26,27,30), but never an RNA-RNA interaction. Therefore, we utilized the specialized capabilities of AUC to provide us with an orthogonal technique that could validate our SEC-MALS results. While both 5'TR and 3'TR are similar in length, the sequences differ significantly,

which will cause a change in secondary structure and, ultimately, tertiary structure. If these potential tertiary structural differences are significant enough, there should be a difference between the 5'TR and 3'TR sedimentation. Both 5'TR and 3'TR show highly monodispersed sedimentation profiles, evident by a single Gaussian distribution. 5'TR is represented by a primary sedimentation profile at ~5.9 S, while 3'TR is represented at ~5.3 S (Figure 9.4C). These distributions confirm that not only are the TRs homogenous, but they also sediment differently based on tertiary structure differences. This difference in sedimentation profiles allowed us to validate our previous SEC-MALS results by mixing 5'TR and 3'TR in a 1:1 ratio and performing another sedimentation velocity experiment. As with previous SEC-MALS experiments, the AUC results show an additional sedimentation peak at ~7.6 S, which is not present in the individual TR experiments allowing us to confirm JEV 5'-3' TR interaction (Figure 9.4C). Additionally, since the concentrations of both RNA were ~200 nM, we can also conclude that the binding affinity (K_D) is likely in the low nanomolar range. We believe this experiment provides the first evidence of an RNA-RNA interaction via AUC, showing that AUC can be an essential tool for showing RNA-RNA interaction while simultaneously determining sample quality and oligomeric state. Taking the AUC and SEC-MALS results together, we can confidently conclude that the additional peak(s) formed in both experiments is the complex of 5'TR and 3'TR, which provides experimental validation of computational predictions.

9.5.4 Determining the RNA-RNA interaction affinity

With the above-mentioned relative binding affinity and direct evidence of interaction, we sought to quantify the affinity of the 5' and 3' TRs using MST. MST allows for studying the interaction between a serially diluted ligand and a fluorescently labeled target by measuring the change in fluorescent migration following excitation by an infrared laser (31). Using the difference between the "cold" and "hot" areas of the MST traces (Figure 9.5A), MST can determine the K_D of the biomolecular interaction. Additionally, Figure 9.5A demonstrates that aggregation of the target was not observed (32) and that monomeric fractions were utilized

in the assay. Our analysis indicates that the TRs interact with a K_D of 60 ± 9 nM (Figure 9.5B), which agrees with the previously reported K_D of 32 ± 1 nM found in WNV using isothermal titration calorimetry (22), which is reasonably similar, validating our findings.

Consequently, the binding affinity further validates our AUC experiment, which was performed with a ~200 nM concentration of JEV TR RNA. To our knowledge, few studies (33,34) have shown evidence of RNA-RNA interactions characterized using MST, highlighting the growing applicability of the technique. Next, we needed to validate whether the CS is the primary driver of this RNA-RNA interaction. We computationally predicted the RNA duplex secondary structure (Figure S9.3), showing numerous potential base-pairing sites across the interaction, but the longest uninterrupted stretch of nucleotides was the CS. We investigated the degree to which the inclusion of the CS impacts binding affinity by exploring various fragments of the JEV TRs. JEV 3'UTR, the full-length 3' non-coding RNA at 573 nt, and JEV 5'TR, which both still include the CS, were assayed and found to interact with a K_D of 169 ± 18 nM (Figure 9.5B). As expected, the theoretically determined highest binding isolates of JEV (5'TR and 3'TR) bound with a slightly higher affinity than the 3'UTR and 5'TR but are within the same magnitude. When excluding the CS through a truncation (JEV 5'UTR-97 nt), the construct lacking 97-224 compared to JEV 5' TR showed that the interaction was almost non-existent with a K_D of 23 ± 6 μM. Finally, we performed the same assay with a mutant version of 3'TR, 3'TR Mut. This construct is identical to JEV 3' TR but has the 11-nucleotides of the CS mutated to no longer base pair to the 5' CS. Therefore, we believed that including the additional long-range binding element would result in a binding tighter than the 5' UTR-3'TR interaction but still considerably weaker than the canonical CS included. The interaction was as expected and was significantly lower than the 5'TR-3'TR interaction, with a K_D of 7 ± 1.4 μM (Figure 9.5B). This change in affinity demonstrates that while the theoretical duplex interaction may have considerable base pairing, the primary driver of the interaction is the cyclization sequence.

9.5.5 RNA-RNA Kinetic and thermodynamic Studies

As observed in Figure 9.3A and Figure 9.4A, both TRs are likely to self-associate even though careful consideration was taken to avoid this. Therefore, we utilized computational analysis to explain the self-association of TRs potentially. This kinetic model builds upon the Turner nearest neighbor energy model for RNA secondary structures (35). While our predictions find potential homodimer interactions with similar thermodynamic stability as the hetero duplex (canonical CS), these homodimer interactions are quite extended and would require extensive refolding. Based on a kinetic analysis (Figure S9.2), we assume that homodimers would form only short, less stable interactions, which would readily dissociate in favour of the more stable heterodimer (3' – 5' complex) when mixed.

To help explain our MST observations, we compared the measured dissociation constants to the predicted interactions for the different TR and UTR constructs. We calculated binding free energies at 37 °C either from the measured dissociation constants or from predictions using RNAup (predictions are in parentheses). For the three constructs, the 5'TR and 3'TR, 5'TR and 3'UTR, and 5'UTR and 3'TR complex, we obtained ΔGs of -10.2 (-10.1) kcal/mol, -9.6 (-7.6) kcal/mol and -6.6 (-7.2) kcal/mol, respectively. The more considerable discrepancy for the second construct (5'TR and 3'UTR) can be explained because the thermodynamic model predicts a refolding (compared to the known consensus structure) when the longer 3'UTR is used, occluding the CS. We also investigated potential additional interaction sites to the known CS. The most promising is a kinetically favourable interaction corresponding to the known upstream AUG Region (UAR) (36) with a predicted stability of -3.6 kcal/mol. This interaction could stabilize the CS interaction but is not strong enough to explain the duplex formation alone. Moreover, in a cellular environment, other factors could likely influence binding kinetics and thermodynamics; however, this evidence shows that this interaction is spontaneous and favorable even in the absence of other cellular factors. One would assume that the contact between the 5' and 3' TRs would be rare in solution due to the

distance between them, but that once a single contact has formed, the rest would spontaneously "zip up." This phenomenon is likely consistent within the cell regarding the amount of (-) sense RNA vs. (+) sense RNA (37). Most (+) sense viral RNA is translated into proteins and not used to replicate the virus.

9.5.6 three-dimensional RNA-RNA interaction analysis through small-angle X-ray scattering

Having characterized the cyclization RNA-RNA binding affinity, we turned to small-angle X-ray scattering to understand the 3-dimensional interaction in solution. Using SEC-SAXS, it is possible to differentiate species based on size and shape prior to actual SAXS measurements, providing increased confidence in monodispersity and allowing for the characterization of a potential complex from two mixed samples (38-41). MST experiments revealed that the binding affinity of the 5'TR and 3' UTR were comparable to the 5'TR-3'TR interaction, so we decided to perform structural experiments on the larger RNA construct (3' UTR) to gain a better characterization of how this RNA interaction is arranged in solution and its relative flexibility.

Three SAXS data sets were collected: 5' TR, 3' UTR, and 5'TR + 3'UTR. Each was merged and represented in figure S9.4A as relative intensity vs. scattering angle. Guinier analysis was performed on each data set to determine the radius of gyration (R_g) for each RNA resulting in 71.61 ± 0.106 Å and 111.5 ± 0.33 Å for 5'TR and 3'UTR respectively (fig S9.4B). Complex (5'TR + 3'UTR) Guinier analysis reveals an R_g of 127.6 ± 0.34 Å, an increase compared to each RNA. The linear regression of each sample demonstrates that each sample is monodispersed and free of any electrostatic interactions between similar molecules (42-44). Intensity data were transformed into dimensionless Kratky analysis data (Fig S9.4C) to evaluate approximate foldedness and conformation (45). Kratky analysis (fig S9.4C) for all three RNA suggests extended conformations of folded RNA based on the relative plateauing of each data set (42). Finally, each data set was evaluated using paired-distance distribution (p(r)) analysis in which reciprocal space data is converted into real-space electron density

data via indirect Fourier transformations (46). P(r) analysis presents real space R_g values of 71.8 ± 0.104 Å, 112.0 ± 0.329 Å, and 128.1 ± 0.355 Å for the 5'TR, 3'UTR, and the complex, respectively (fig S9.4D). Real-space R_g values have a very high level of agreement with the reciprocal-space values, indicating the validity of the entire data set. Furthermore, the shape of each P(r) plot indicates an elongated conformation, whereas a gaussian distribution with a peak at $D_{max}/2$ would be indicative of a spherical, globular molecule. The maximum distance for each RNA data set was 224 Å, 345 Å, and 400 Å for the 5'TR, 3' UTR, and complex, respectively. Notably, the D_{max} values illustrate that while the complex is larger than either individual RNA, likely that the interaction is not end-to-end. The maximal distance for an end-to-end interaction would be considerably larger than 400 Å, suggesting there may be considerable overlap between the 5'TR and 3'UTR.

Ab inito modeling through DAMMIN was then performed on 5'TR and 3'UTR data sets to generate low-resolution three-dimensional structures. 100 models were generated for each RNA showing favorable agreement via χ^2 values of 1.15 and 1.12 for the 5'TR and 3'UTR, respectively. The 5' TR was filtered and merged into a singular representative structure with a normalized spatial discrepancy (NSD) of 0.937 ± 0.020, indicating a good fit of each structure to the representative filtered model (Figure 9.7A)(47). Represented in blue, the 5' TR is an elongated RNA structure like other RNA of similar size (27,30,48). The 3'UTR, however, could not be filtered and averaged together to get a singular representative model, even though the average χ^2 value was 1.12, suggesting that each model strongly agreed with the original scattering data. We, therefore, chose to cluster the 100 respective models from the 3'UTR, resulting in 22 distinctly different clusters represented by the pie chart in Figure 9.7B. This suggests that the 3'UTR has flexible regions, similar to other long non-coding RNA (49,50). The three largest clusters of models are color coded to the pie chart, and while each is classified as a distinct structure, they contain a similar overall architecture: each having a large center pocket and large portions of electron density towards each terminus.

Importantly, since the affinity of the 5'TR – 3'UTR interaction was sufficiently low, we believed that the complex would remain intact throughout SEC-SAXS, allowing for peak separation and, ultimately, data collection. Creation of the three-dimensional models required multiple distinct inputs, differing from the terminal regions individually. Using MONSA, multi-phase bead modeling can simultaneously fit multiple scattering curves to a single data set (51,52). Therefore, discerning which portion of the total scattering data was contributed by each RNA element required inputting data collected from the 5'TR, the 3'TR, and the complex. Importantly, when evaluating χ^2 values for the complex, each input is scored based on its agreement with the raw scattering data with approximate values of 1.20, 1.40, and 1.32 for the 5'TR, 3'UTR, and complex, respectively. These values show that each distinct scattering curve fits into the data set and has a good agreement. Like the 3'UTR on its own, a singular averaged model cannot accurately represent the RNA, which led us to cluster the model representations (Figure 9.7C) similarly. This clustering resulted in 11 distinct representations of the solution structure, half as many clusters as the 3'UTR alone, with 66% of the total structures falling into 3 clusters. Clustering provides evidence that the interaction causes a change in the conformational space and a subsequent reduction in the flexibility of the 3'UTR. Figure 9.7C represents the top 3 conformational clusters, with the 5'TR in a consistent blue color and the 3'UTR in pink/red. The 5'TR remains relatively unchanged upon interaction with the 3'UTR, still adopting an elongated shape similar to its previous conformation (Figure 9.7A). Most of the conformational change to the complex is changes to the 3'UTR. Overall, the central pocket, which was visible in the top 3'UTR conformations, is no longer visible, and the electron density has been re-distributed from being concentrated around the ends of the molecule (see Figure 9.7B). Furthermore, the 3-dimensional complex shows that there is likely not considerable 5'-3' duplex formation, with generally 1-2 contact points, one of which is likely the CS. Structural information about this complex can also inform on the potential mechanism by which the flaviviral RNA-dependent RNA polymerase NS5 interacts with the 5' TR SLA yet transcribes the negative sense genome 3' to 5' (14,53-55). However, the nature of SAXS

being low-resolution means we cannot accurately pinpoint specific RNA motifs such as the CS or the 3’ or 5’ ends. Future directions will need to be focused on gaining a higher-resolution model of the interaction through either computational modeling with SAXS envelopes or cryo-EM imaging.

9.6 Conclusion

Our study demonstrates that combining computational and biophysical approaches can provide detailed insights into RNA-RNA interactions that govern many fundamental biological processes. Our SEC-MALS and AUC data also suggest, for the first time, that these techniques can be used to study RNA-RNA interactions in solution while confirming that MST is a powerful technique for affinity determination of RNA-RNA interactions. We demonstrate that computational analysis, such as kinetic landscapes, can provide vital evidence for predicting RNA-RNA interactions and even sites of interactions. Additionally, our binding affinity data presents that the CS is the primary driver of terminal region interaction with a nanomolar affinity. This binding affinity is critical to developing potential inhibitory therapeutics targeting the cyclization sequence. Additionally, we present the 3-dimensional low-resolution structure complex structure in solution, providing evidence that it is likely flexible and dynamic, with a small amount of duplex formation between the 5’ TR and 3’UTR. Our work directly contributes to understanding the Flaviviral cyclization conservation that could help develop therapeutics with a potential for multi-virus inhibition. Moreover, we believe our comprehensive computational and biophysical pipeline can be applied to virtually any RNA-RNA interacting system in solution so long as proper sample purification and quality control are forefronts. We hope this study may influence researchers to further investigate the importance of the cyclization sequence of JEV and other flaviviruses such as ZIKV and DENV *in vivo*. Additionally, *in vivo* experiments investigating sequence conservation of the CS could be instrumental to understanding the evolutionary relationship(s) of this family of viruses.

9.7 Materials and Methods

9.7.1 Computational assessment of RNA-RNA interaction genome cyclization

Putative interaction sites between 3' and 5'TRs of 109 JEV isolates were predicted with IntaRNA v3.2.0, RNAup v2.4.18, and RNAcofold (56-60). A consensus secondary structure of the 5'/ 3' terminal regions in 20 phylogenetically related flaviviruses was computed with RNAalifold v2.4.18 (61) from the ViennaRNA Package v2.4.18 (62), based on a structural nucleotide multiple sequence alignment computed with LocARNA v2.0.0RC8 (63).

We evaluated whether the predicted TR interactions are kinetically feasible by applying a novel direct path model. In this model, the full target interaction is formed starting from an initial seed interaction, which folds into the target interaction by adding or removing target base pairs. All substructures on a folding path consist of consecutive base pairs of the target interaction. Consequently, paths are 'direct' since they cannot contain detours via non-target interaction base pairs. We compute the free energy of each substructure analogous to the RNAup energy model. The two main contributions are the cost of making the interaction site unpaired in the two intramolecular structures and the stabilizing contribution of the interaction base pairs. Both contributions were obtained from the ViennaRNA library. We can model interaction formation as a Markov process based on this direct paths model and the energy model/function. The overall rate at which an interaction is formed is determined by the energy barrier along the direct folding path (activation energy). Since the model's possible substructures (aka states) can be completely described by their first base pair *i* and last base pair *j*, they form a 2D energy landscape, and any direct path from a first base pair to the full interaction can be drawn, as shown in figure S9.2. Plots of these energy landscapes allow for the visual identification of barriers along the folding path.

9.7.2 Preparation and Purification of non-coding RNA

JEV 3'TR, 3' TR Mut, and 5'TR constructs of JEV were designed based on the GenBank sequence of KR265316.1, while JEV 5' and 3' UTR were based on KT957419.1. cDNA sequences were prepared in pUC57 plasmids under the control of T7 RNA polymerase. The cDNA was flanked by two additional G nucleotides on the 5' end and an XbaI restriction enzyme cut site (T^CTAGA) on the 3' end. The 3'TR and 5'TR constructs are graphically represented in Figure 9.1A to enhance clarity on which portion of the terminal region is involved in each experiment. All RNA construct sequences are listed below with an underlined region representing the theoretical cyclization sequence (CS) which forms the 11-nucleotide base pairing complement between the 5' and 3'TR. Briefly, JEV 5' UTR serves as a cyclization sequence truncation mutant, having nucleotides 97-224 removed compared to JEV 5' TR. JEV 3' TR Mut is an exact copy of JEV 3' TR, with only the 11-nucleotide cyclization sequence mutated to no longer base pair to JEV 5' TR.

Table 9.1. Construct information for RNA constructs used.

Name	**Genomic Location**	**Length (NTs)**	**Sequence (5' – 3')**	**Mw (kDa)**
JEV 5' TR	1-224	228	GGAGAAGUUUAUCUGUGUGAACUUCUUGGCUUAG UAUCGUUGAGAAGAAUCGAGAGAUUAGUGCAGUU UAAACAGUUUUUUAGAACGGAAGAUAACCAUGACU AAAAAACCAGGAGGGCCCGGUAAAAACCGGGCUAU CAAUAUGCUGAAACGCGGCUUACCCCGCGUAUUCC CACUAGUGGGAGUGAAGAGGGUAGUAAUGAGCUU GUUGGACGGCAGAGGGCCAGU	73.9
JEV 5' UTR	1-96	100	GGAGAAGUUUUAUCGUGUGAACUUCUUGGCUUAG UAUCGUUGAGAAGAAUCGAGAGAUUAGUGCAGUU UAAACAGUUUUUUAGAACGGAAGAACAACCU	32.2
JEV 3' TR	10758-10976	221	GGUUAGAGGAGACCCCGUGGAAACAACAACAUGC GGCCCAAGCCCCCUCGAAGCUGUAUAGGAGGUGG AAGGACUAGAGGUUAGAGGAGACCCCGCAUUUGC AUCAAACAGCAUAUUGACACCUGGGAAUAGACUAG GAGAUCUUCUGCUCUAUCUCAACAUCACUUACUAG GCACAUAUCGCCGAAGUAUGUAGCUGGUGGUAGG AAGAACACACGAUCU	71.4
JEV 3' TR Mut	10758-10976	221	GGUUAGAGGAGACCCCGUGGAAACAACAACAUGC GGCCCAAGCCCCCUCGAAGCUGUAUAGGAGGUGG AAGGACUAGAGGUUAGAGGAGACCCCGCAUUUGC AUCAAAAUCUUCGAACUCACCUGGGAAUAGACUAG GAGAUCUUCUGCUCUAUCUCAACAUCACUUACUAG GCACAUAUCGCCGAAGUAUGUAGCUGGUGGUAGG AAGAACACACGAUCU	71.4

JEV 3’ UTR	10395-10965	573	GGACAGGAUAAAGUCAUAUGUGUAAUGUGAGAUA AGAAAAUGUGCAUGUGGAGUCAGGCCAGCAAAAG CUGCCACCGGAUACUGAGUAGACGGUGCUGCCUG CGUCUCAGUCCCAGGAGGACUGGGUUAACAAAUC UGACAACGGAAGGUGGGAAAGCCCUCAGAACCGU CUCGGAAGCAGGUCCCUGCUCACCGGAAGUUGAA AGACCAACGUCAGGCCACAAUUCUGUGCCACUCCG CUGGGGAGUGCGGCCUGCGCAGCCCCAGGAGGAC UGGGUUAACAAAGCCGUUGAGGCCCCCACGGCCC AAGCCUCGUCUAAGAUGCAAUAGACUAGGUGUAAG GACUAGAGGUUAGAGGAGACCCCGUGGAAACAAC AUUGUGCGGCCCAAACCCCCUCGAAGCUGUAGAG GAGGUGGAAGGACUAGAGGUUAGAGGAGACCCCG CAUUUGCAUCAAAACAGCAUAUUGACACCUGGGAA UAGACUGGGAGAUCUUCUGCUCUAUCUCAACAUCA GCUACUAGGCACAGAGCGCCGAAGUAUGUAGCUG GUGGUGAGGAAGAAACCAGGAUCUU	185.6

RNA was prepared via *in vitro* transcription reaction using T7 RNA polymerase (purified in-house) followed by size-exclusion chromatography (SEC) purification using a Superdex 200 Increase GL 10/300 (Global Life Science Solutions USA LLC, Marlborough, MA, USA) in JEV RNA Buffer (10 mM Bis-tris pH 5.0, 100 mM NaCl, 15 mM KCl 15 mM $MgCl_2$, 10% glycerol) via an ÄKTA pure FPLC (Global Life Science Solutions USA LLC, Marlborough, MA, USA) with a flow rate of 0.5 mL/min. Urea-polyacrylamide gel electrophoresis (Urea-PAGE) was utilized to analyze SEC peak fractions. We mixed 10 µL of each fraction with 2 µL of denaturing RNA loading dye and loaded it into a 1.0 cm well PAGE casting plate (Bio-Rad Laboratories, Mississauga, ON, Canada). Urea-PAGE (7.5%) was run at room temperature, 300V, for 25 min (20 min for JEV 5' TR) in 0.5x TBE (Tris-Borate-EDTA) buffer (heated), followed by staining with Sybr safe (Thermofisher Scientific, Saint-Laurant, QC, Canada) and visualization. Fractions containing a single band were deemed acceptable and used in subsequent experiments.

9.7.3 Light Scattering

Multi-angle light scattering (MALS) experiments were performed on a Dawn® (Wyatt Technology Corporation, Santa Barbara, CA, USA) multi-angle light scattering instrument with 18 detector angles utilizing a 658 nm laser. Furthermore, an Optilab® (Wyatt Technology Corporation, Santa Barbara, CA, USA) refractometer was also positioned downstream to

measure the solvent refractive index and absolute concentration of solutes. These instruments were positioned in line with an SEC column (Superdex 200 increase 10/300 GL, Global Life Science Solutions, USA LLC, Marlborough, MA, USA) attached to an ÄKTA pure FPLC (SEC-MALS). All experiments were performed at ambient room temperature (20°C) with the same flow rate and buffer as previous SEC experiments described above. The refractive index of the solvent was defined as 1.3308 (measured by in-line Optilab® refractometer), while the dn/dc (refractive index increment) value of 0.1720 mL/g was used for all RNAs (64). The final concentration of 3′ TR and 5′ TR RNA used was 1.3 μM and 3.5 μM, respectively, in a combined volume of 500 μL. Samples were incubated for 3 hours prior to loading. Data were analyzed using Astra v8.0.0.25, and absolute molecular weight (M_w) was calculated using Equation 1 for each elution point, where: $R(\theta)$ is Rayleigh's ratio, K is the polymer constant, and c is the concentration of the solution.

$$M_w = \frac{R(\theta)}{K^* c} \tag{9.1}$$

9.7.4 Analytical Ultra-Centrifugation

AUC data for 5′ TR and 3′ TR were collected using a Beckman Optima AUC centrifuge and an AN50-Ti rotor at 20°C at the Canadian Centre for Hydrodynamics at the University of Lethbridge. We measured the 5′TR (214 nM, 0.5 OD @260 nm), 3′TR (224 nM, 0.5 OD @260 nm), and a 1:1 mixture of 5′TR and 3′TR samples into standard Beckman Coulter cell housings equipped with Epon-2 channel centerpieces, and fitted with sapphire windows, in JEV RNA Buffer. As a first step, we centrifuged samples at 25,000 rpm and collected scans at 20-second intervals. We used the UltraScan-III package (65) to analyze all data via supercomputer calculations in-house. We analyzed the SV-AUC data using two-dimensional spectrum analysis (2DSA) with simultaneous removal of time-invariant noise, meniscus, and

bottom positions fitted, followed by enhanced van Holde-Weischet analysis (66). We estimated the buffer density and viscosity corrections with UltraScan (1.0269g/cm3 and 1.293 cP, respectively). All hydrodynamic parameters were corrected to standard conditions at 20°C and water

9.7.5 Fluorescent labeling of RNA

Purified RNA was subjected to labeling at the 5’ end by the fluorophore Alexa 488 (A488) (Thermofisher Scientific, Saint-Laurant, QC, Canada). One milligram of A488 was resuspended in 175µL of 0.2M KCl. 7.5µL concentrated RNA (>100µM) was added to 1.25mg 1-ethyl-3-(3-dimethylamino) propyl carbodiimide hydrochloride (EDC) prior to the addition of 10µL resuspended A488. Samples were vortexed until contents were dissolved entirely before adding 20µL 0.1M imidazole, pH 6. Reactions were incubated in a 37°C water bath overnight in the absence of light, followed by the removal of free dye through 10 kDa Vivaspin® 500 centrifugal concentrators (Sartorius Stedim Biotech, Göttingen, Lower Saxony, Germany). After removing all free dye, labeled RNA was diluted in JEV RNA Buffer, and fluorescence checks were conducted using microscale thermophoresis.

9.7.6 Microscale thermophoresis

A 2-fold serial dilution was performed on the RNA ligand, either 3’TR, 3’TR Mut, or 3’UTR, where the highest concentration in the assay was 21 µM, 9.8 µM, and 5.5 µM, respectively. A constant amount of fluorescently labeled RNA Target, 5’TR, or 5’UTR, was added to each serial dilution of RNA ligand resulting in a final concentration of 25 nM and 52 nM, respectively. Mixtures were incubated at room temperature for 3 hours and then loaded into a Nanotemper Technologies Monolith® NT.115 instrument (Munich, Germany) using standard capillaries. Thermophoresis was measured at room temperature (22°C) and performed using 100% excitation power along with medium IR-Laser power. Initial fluorescence migration was measured from (-1.0 to 0 s) and used to normalize the measured

fluorescent migration time (9.0 to 10.0 s). Three independent replicates were analyzed using MO.Affinity Analysis software v2.1.3 and fit to the standard K_D fit model, which describes a 1:1 stoichiometric molecular interaction according to the law of mass action. The dissociation constant (K_D) is estimated by fitting equation 1., where *F(c)* is the fraction bound at a given ligand concentration c. Unbound is the F_{norm} signal of the isolated target; Bound is the F_{norm} signal of the complex, while c_{target} is the final concentration of the target in the specific assay.

$$\boldsymbol{F(c) = Unbound + (Bound - Unbound) \times \frac{c + c_{target} + K_d - \sqrt{(c + c_{target} + K_d)^2 - 4\, c\, c_{target}}}{2\, c_{target}}} \quad (9.2)$$

9.7.7 Small-angle X-ray scattering

RNA sample data collection was performed on the B21 HPLC-SAXS beamline at Diamond Light Source (Didcot, Oxfordshire, UK), as reported elsewhere (67). An Agilent 1200 (Agilent Technologies, Stockport, UK) HPLC was utilized through connection to a specialized flow-cell, whereas 50 µL of each purified RNA (5′ TR, 3′ UTR, and 5′ TR+3′UTR respectively) were injected into a JEV RNA buffer equilibrated Shodex 403KW-4F HPLC column (Showa Denko America Inc., New York, NY, USA) with a flow rate of 0.160 mL/min. Concentrations were 1.2 mg/mL and 1.1 mg/mL for 5′ TR and 3′ UTR, respectively, and the complex was a 1:1 volume mixture of both. Frames were exposed to synchrotron radiation (x-rays) for 3 seconds for a total of ~600 frames. The resulting data was buffer subtracted using Chromixs (68) for each sample peak, and then data analysis was performed using the ATSAS suite of programs (69). The radius of gyration (R_g) was evaluated through Guinier analysis, while additionally determining sample quality (70) and relative foldedness of RNA molecules was determined via dimensionless Kratky analysis (71). GNOM was used to perform paired distance distribution P(r) analysis to obtain real space Rg and maximum particle dimension (D_{max}) measurements (72,73). Using P(r) derived information, 20 models were generated for 5′ TR and 100 models for 3′ UTR using DAMMIN (51). Following simulated annealing via

DAMMIN, 5′ TR models were averaged and filtered to produce a single representative model using DAMAVER and DAMFILT (51,74). Given the size of 3′ UTR, we decided to cluster representative models instead of generating a singular averaged model via DAMCLUST (74). Model reconstruction for the 5′ TR and 3′ UTR complex was generated through MONSA (51) using data input from 5′TR, 3′UTR, and data from 5′TR+3′UTR. 100 models were generated and then partitioned into representative clusters via DAMCLUST.

9.8 Supplementary Information

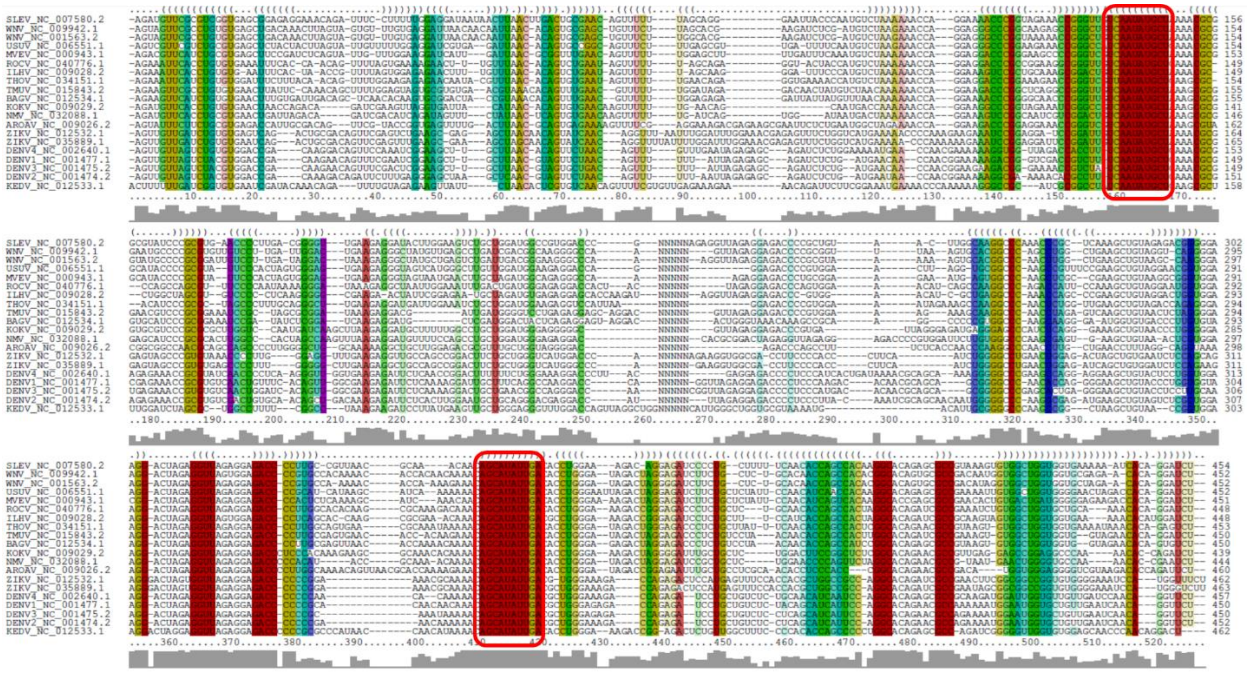


Figure S9.1. Multiple sequence alignment for 5'TR and 3'TR of 20 mosquito-borne Flaviviruses that was used to form the consensus structure in Fig. 9.2. The alignment was obtained from the structural alignment toll LocARNA, the two fragments were joined by a linker of 5 Ns. Coloring indicates covariation in consensus base pairs following the RNAalifold schema, ranging from red (no covariation, full primary sequence conservation) to violet (full covariation, all six possible combinations of base pairs). Red boxes highlight the conserved 5' and 3' cyclization sequence, respectively.

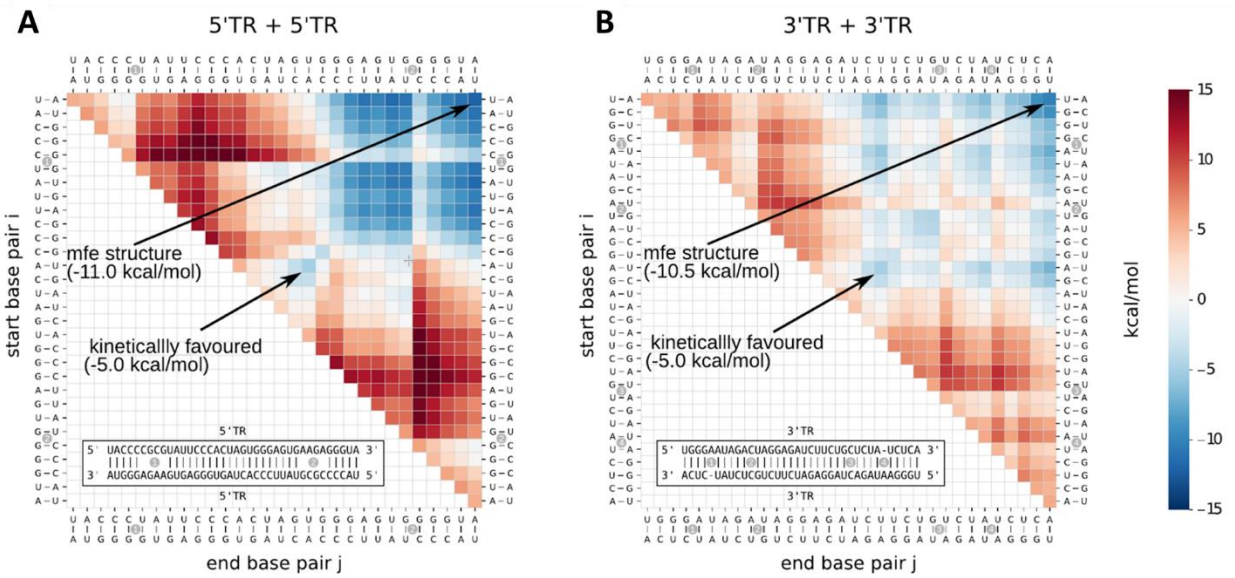


Figure S9.2. Energy landscapes of predicted homodimer interactions. A) 5'TR homodimer **B)** 3'TR homodimer. The thermodynamical model predicts stable minimum free energy structures with -11 kcal/mol and -10.5 kcal/mol, for the two respective homodimers. Based on the energy landscapes, there is no kinetically favourable path from a first base pair (drawn on the diagonal) to the full interaction (drawn in the upper right corner) through which the full interaction could be reached without crossing any energy barriers. In both examples, the most stable substructure that is energetically favourable has a stability of -5 kcal/mol and therefore is less stable than the known CS.

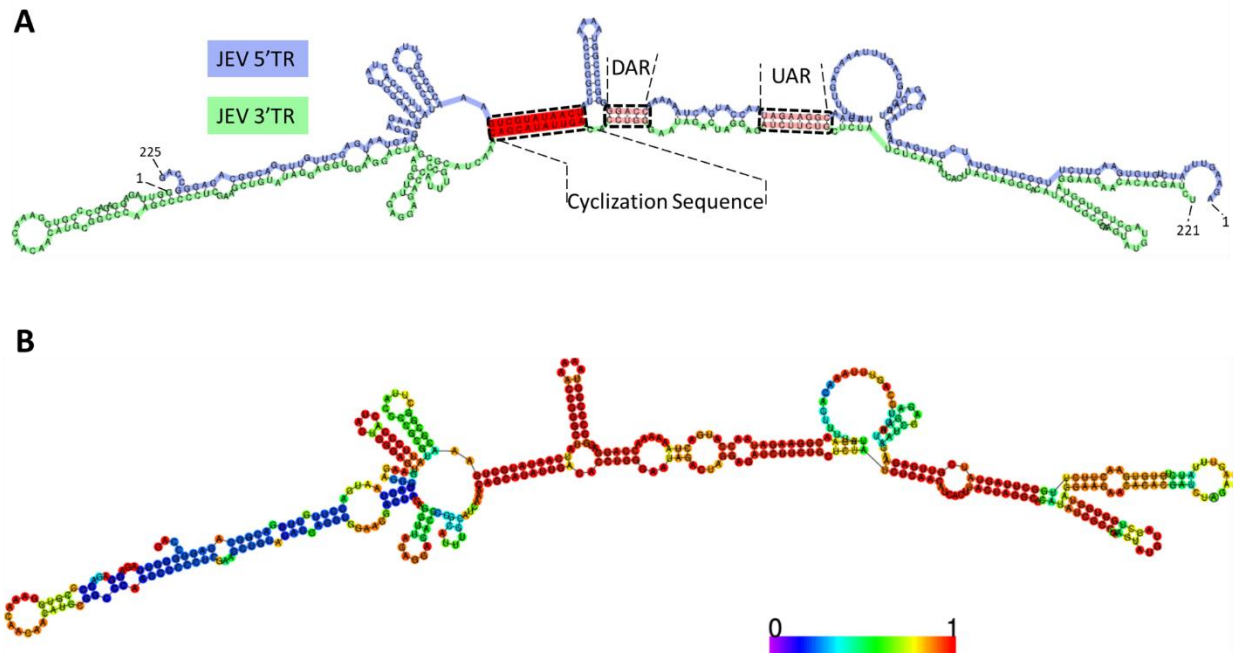


Figure S9.3. Predicted duplex structure of the JEV 3'TR and JEV 5'TR obtained from RNAcofold, showing the maximum possible extent of interaction. A) Known long range interaction elements, such as cyclization sequence (CS), downstream AUG region (DAR) and upstream AUG region (UAR) are highlighted. **B)** structure with reliability annotation (based on pair probabilities) ranging from blue (poor reliability) to red (high confidence).

9.9 Data availability

We selected complete JEV isolates (including 3′ and 5′ UTR) from NCBI GenBank and used the first 225 nts (5′TR) and the last 221 nts (3′TR) to perform interaction predictions. The corresponding accession numbers are included in the supplementary information. SAXS data has been deposited to SASDB under accession numbers SASDQL9, SASDQM9, and SASDQN9 for JEV 5′ TR, JEV 3′UTR, and JEV 5′TR+3′UTR complex, respectively.

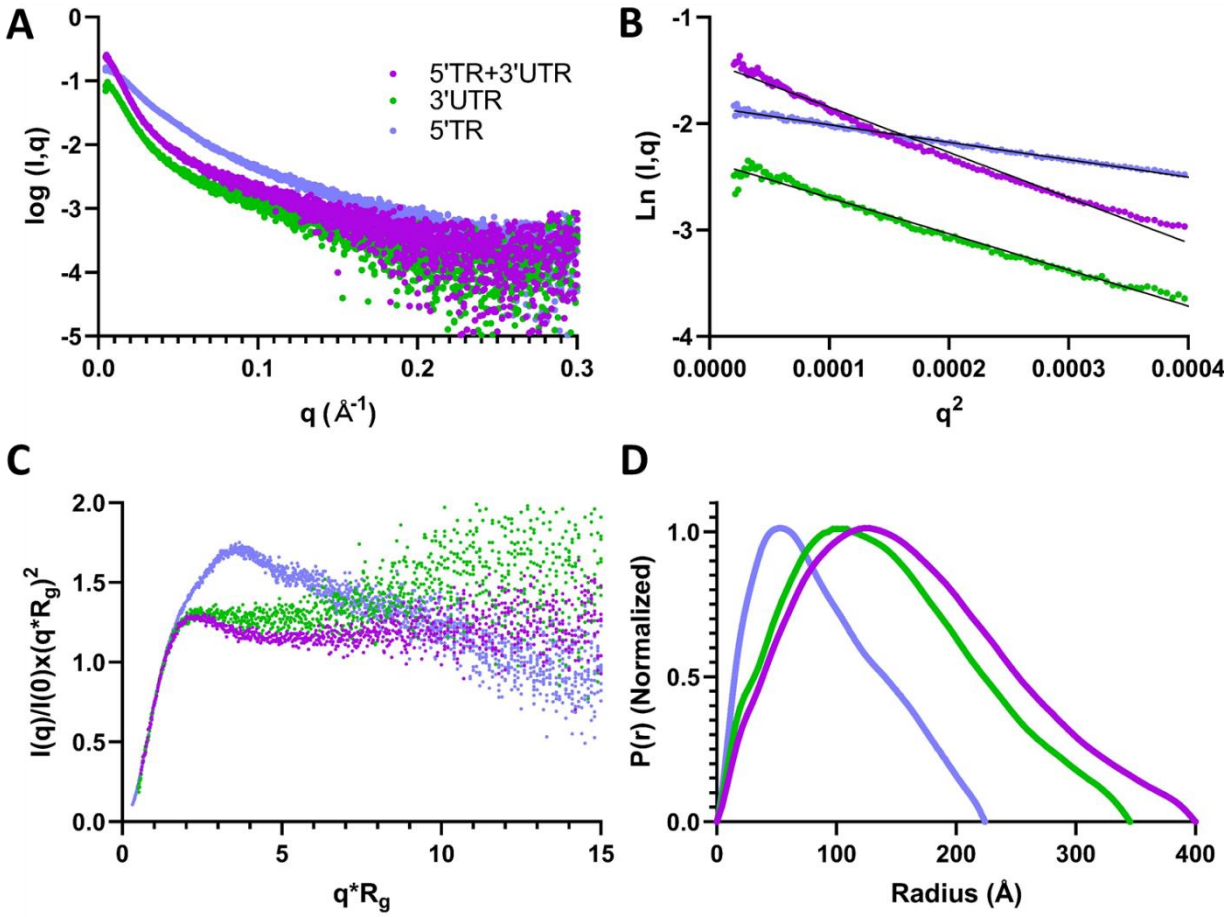


Figure S9.4. Small Angle X-Ray Scattering (SAXS) plots of JEV RNA-RNA interaction. A) Subtracted & merged scattering data for JEV RNA depicting the scattering intensity vs. scattering angle ($q = 4\pi\sin\Theta/\lambda$). **B)** Guinier plots representing the determination of R_g and homogeneity derived from the low-angle region. **C)** Dimensionless Kratky plots of JEV RNA depicting an elongated structure as a result of the non-Gaussian shape(s) of the curve(s). **D)** Pair-distance distribution plots for JEV RNA for real-space R_g and maximal particle dimension (D_{max}) determination from the entire SAXS dataset.

9.10 References

1. Zhang, F. and Lupski, J.R. (2015) Non-coding genetic variants in human disease. *Hum Mol Genet*, **24**, R102-110.
2. Balas, M.M., Hartwick, E.W., Barrington, C., Roberts, J.T., Wu, S.K., Bettcher, R., Griffin, A.M., Kieft, J.S. and Johnson, A.M. (2021) Establishing RNA-RNA interactions remodels lncRNA structure and promotes PRC2 activity. *Science Advances*, **7**, eabc9191.
3. Aguilar, R., Spencer, K.B., Kesner, B., Rizvi, N.F., Badmalia, M.D., Mrozowich, T., Mortison, J.D., Rivera, C., Smith, G.F., Burchard, J. *et al.* (2022) Targeting Xist with compounds that disrupt RNA structure and X inactivation. *Nature*, **604**, 160-166.
4. Jain, A. and Vale, R.D. (2017) RNA phase transitions in repeat expansion disorders. *Nature*, **546**, 243-247.
5. Van Treeck, B. and Parker, R. (2018) Emerging Roles for Intermolecular RNA-RNA Interactions in RNP Assemblies. *Cell*, **174**, 791-802.
6. Heffelfinger, J.D., Li, X., Batmunkh, N., Grabovac, V., Diorditsa, S., Liyanage, J.B., Pattamadilok, S., Bahl, S., Vannice, K.S., Hyde, T.B. *et al.* (2017) Japanese Encephalitis Surveillance and Immunization - Asia and Western Pacific Regions, 2016. *MMWR Morb Mortal Wkly Rep*, **66**, 579-583.
7. Wang, Q.-Y. and Shi, P.-Y. (2015) Flavivirus Entry Inhibitors. *ACS Infectious Diseases*, **1**, 428-434.
8. Huang, Y.J., Higgs, S., Horne, K.M. and Vanlandingham, D.L. (2014) Flavivirus-mosquito interactions. *Viruses*, **6**, 4703-4730.
9. Tjaden, N.B., Caminade, C., Beierkuhnlein, C. and Thomas, S.M. (2018) Mosquito-Borne Diseases: Advances in Modelling Climate-Change Impacts. *Trends in Parasitology*, **34**, 227-245.
10. Lindenbach, B., Thiel, H.J. and Rice, C.M. (2007) Flaviviridae: The viruses and their replication. *Fields Virology*, 1101-1151.
11. Ng, W.C., Soto-Acosta, R., Bradrick, S.S., Garcia-Blanco, M.A. and Ooi, E.E. (2017) The 5′ and 3′ Untranslated Regions of the Flaviviral Genome. *Viruses*, **9**, 137.
12. Bollati, M., Alvarez, K., Assenberg, R., Baronti, C., Canard, B., Cook, S., Coutard, B., Decroly, E., de Lamballerie, X., Gould, E.A. *et al.* (2010) Structure and functionality in flavivirus NS-proteins: perspectives for drug design. *Antiviral research*, **87**, 125-148.
13. Mandl, C.W., Holzmann, H., Kunz, C. and Heinz, F.X. (1993) Complete Genomic Sequence of Powassan Virus: Evaluation of Genetic Elements in Tick-Borne versus Mosquito-Borne Flaviviruses. *Virology*, **194**, 173-184.
14. Filomatori, C.V., Lodeiro, M.F., Alvarez, D.E., Samsa, M.M., Pietrasanta, L. and Gamarnik, A.V. (2006) A 5' RNA element promotes dengue virus RNA synthesis on a circular genome. *Genes Dev*, **20**, 2238-2249.
15. You, S. and Padmanabhan, R. (1999) A Novel in Vitro Replication System for Dengue Virus: INITIATION OF RNA SYNTHESIS AT THE 3′-END OF EXOGENOUS VIRAL RNA TEMPLATES REQUIRES 5′- AND 3′-TERMINAL COMPLEMENTARY SEQUENCE MOTIFS OF THE VIRAL RNA*. *Journal of Biological Chemistry*, **274**, 33714-33722.
16. Zhang, B., Dong, H., Stein, D.A., Iversen, P.L. and Shi, P.-Y. (2008) West Nile virus genome cyclization and RNA replication require two pairs of long-distance RNA interactions. *Virology*, **373**, 1-13.
17. Alvarez, D.E., Lodeiro, M.F., Ludueña, S.J., Pietrasanta, L.I. and Gamarnik, A.V. (2005) Long-range RNA-RNA interactions circularize the dengue virus genome. *Journal of virology*, **79**, 6631-6643.

18. Hahn, C.S., Hahn, Y.S., Rice, C.M., Lee, E., Dalgarno, L., Strauss, E.G. and Strauss, J.H. (1987) Conserved elements in the 3' untranslated region of flavivirus RNAs and potential cyclization sequences. *J Mol Biol*, **198**, 33-41.
19. Meier-Stephenson, V., Mrozowich, T., Pham, M. and Patel, T.R. (2018) DEAD-box helicases: the Yin and Yang roles in viral infections. *Biotechnology & genetic engineering reviews*, **34**, 3-32.
20. Nelson, C., Mrozowich, T., Gemmill, D.L., Park, S.M. and Patel, T.R. (2021) Human DDX3X Unwinds Japanese Encephalitis and Zika Viral 5' Terminal Regions. *Int J Mol Sci*, **22**, 413.
21. Men, R., Bray, M., Clark, D., Chanock, R.M. and Lai, C.J. (1996) Dengue type 4 virus mutants containing deletions in the 3' noncoding region of the RNA genome: analysis of growth restriction in cell culture and altered viremia pattern and immunogenicity in rhesus monkeys. *Journal of virology*, **70**, 3930-3937.
22. Deo, S., Patel, T.R., Chojnowski, G., Koul, A., Dzananovic, E., McEleney, K., Bujnicki, J.M. and McKenna, S.A. (2015) Characterization of the termini of the West Nile virus genome and their interactions with the small isoform of the 2' 5'-oligoadenylate synthetase family. *Journal of structural biology*, **190**, 236-249.
23. Thurner, C., Witwer, C., Hofacker, I.L. and Stadler, P.F. (2004) Conserved RNA secondary structures in Flaviviridae genomes. *Journal of General Virology*, **85**, 1113-1124.
24. Ochsenreiter, R., Hofacker, I.L. and Wolfinger, M.T. (2019) Functional RNA Structures in the 3'UTR of Tick-Borne, Insect-Specific and No-Known-Vector Flaviviruses. *Viruses*, **11**, 298.
25. Michael T. Wolfinger, R.O., Ivo L. Hofacker. (2021) In Marz, D. F. a. M. (ed.), *Virus Bioinformatics*. 1 ed. Chapman and Hall/CRC, pp. 36.
26. Nelson, C.R., Mrozowich, T., Park, S.M., D'souza, S., Henrickson, A., Vigar, J.R.J., Wieden, H.-J., Owens, R.J., Demeler, B. and Patel, T.R. (2021) Human DDX17 Unwinds Rift Valley Fever Virus Non-Coding RNAs. *Int J Mol Sci*, **22**, 54.
27. Mrozowich, T., Henrickson, A., Demeler, B. and Patel, T.R. (2020) Nanoscale Structure Determination of Murray Valley Encephalitis and Powassan Virus Non-Coding RNAs. *Viruses*, **12**, 190.
28. Wyatt, P.J. (1993) Light scattering and the absolute characterization of macromolecules. *Analytica Chimica Acta*, **272**, 1-40.
29. Patel, T.R., Winzor, D.J. and Scott, D.J. (2016) Analytical ultracentrifugation: A versatile tool for the characterisation of macromolecular complexes in solution. *Methods*, **95**, 55-61.
30. D'Souza, M.H., Mrozowich, T., Badmalia, M.D., Geeraert, M., Frederickson, A., Henrickson, A., Demeler, B., Wolfinger, M.T. and Patel, T.R. (2022) Biophysical characterisation of human LincRNA-p21 sense and antisense Alu inverted repeats. *Nucleic acids research*, **50**, 5881-5898.
31. Wienken, C.J., Baaske, P., Rothbauer, U., Braun, D. and Duhr, S. (2010) Protein-binding assays in biological liquids using microscale thermophoresis. *Nature Communications*, **1**, 100.
32. Mrozowich, T., MeierStephenson, V. and Patel, T.R. (2019) Microscale thermophoresis: warming up to a new biomolecular interaction technique. *The Biochemist*, **41**, 8-12.
33. Swain, M., Ageeli, A.A., Kasprzak, W.K., Li, M., Miller, J.T., Sztuba-Solinska, J., Schneekloth, J.S., Koirala, D., Piccirili, J., Fraboni, A.J. *et al.* (2021) Dynamic bulge nucleotides in the KSHV PAN ENE triple helix provide a unique binding platform for small molecule ligands. *Nucleic acids research*, **49**, 13179-13193.
34. Jordan, B., Nickel, L. and Schmitz, R.A. (2022), *Prokaryotic Gene Regulation: Methods and Protocols*. Springer, pp. 291-303.
35. Turner, D.H. and Mathews, D.H. (2010) NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic acids research*, **38**, D280-282.
36. Friebe, P., Shi, P.-Y. and Harris, E. (2011) The 5' and 3' downstream AUG region elements are required for mosquito-borne flavivirus RNA replication. *Journal of virology*, **85**, 1900-1905.

37. Chu, P.W.G. and Westaway, E.G. (1985) Replication strategy of Kunjin Virus: Evidence for recycling role of replicative form RNA as template in semiconservative and asymmetric replication. *Virology*, **140**, 68-79.
38. Brosey, C.A. and Tainer, J.A. (2019) Evolving SAXS versatility: solution X-ray scattering for macromolecular architecture, functional landscapes, and integrative structural biology. *Curr Opin Struct Biol*, **58**, 197-213.
39. Pérez, J. and Vachette, P. (2017) In Chaudhuri, B., Muñoz, I. G., Qian, S. and Urban, V. S. (eds.), *Biological Small Angle Scattering: Techniques, Strategies and Tips*. Springer Singapore, Singapore, pp. 183-199.
40. O'Brien, D.P., Brier, S., Ladant, D., Durand, D., Chenal, A. and Vachette, P. (2018) SEC-SAXS and HDX-MS: A powerful combination. The case of the calcium-binding domain of a bacterial toxin. *Biotechnol Appl Biochem*, **65**, 62-68.
41. Graewert, M.A., Da Vela, S., Gräwert, T.W., Molodenskiy, D.S., Blanchet, C.E., Svergun, D.I. and Jeffries, C.M. (2020) Adding Size Exclusion Chromatography (SEC) and Light Scattering (LS) Devices to Obtain High-Quality Small Angle X-Ray Scattering (SAXS) Data. *Crystals*, **10**, 975.
42. Putnam, C.D., Hammel, M., Hura, G.L. and Tainer, J.A. (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys*, **40**, 191-285.
43. Grant, T.D., Luft, J.R., Carter, L.G., Matsui, T., Weiss, T.M., Martel, A. and Snell, E.H. (2015) The accurate assessment of small-angle X-ray scattering data. *Acta Crystallogr D Biol Crystallogr*, **71**, 45-56.
44. Cantara, W.A., Olson, E.D. and Musier-Forsyth, K. (2017) Analysis of RNA structure using small-angle X-ray scattering. *Methods*, **113**, 46-55.
45. Rambo, R.P. and Tainer, J.A. (2011) Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers*, **95**, 559-571.
46. Kikhney, A.G. and Svergun, D.I. (2015) A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett*, **589**, 2570-2577.
47. Kozin, M.B., Svergun, D.I. and Embl, B. (2001) Automated matching of high- and low-resolution structural models. *J. Appl. Cryst*, **34**, 33-41.
48. Jones, C.P., Cantara, W.A., Olson, E.D. and Musier-Forsyth, K. (2014) Small-angle X-ray scattering-derived structure of the HIV-1 5' UTR reveals 3D tRNA mimicry. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 3395-3400.
49. Spokoini-Stern, R., Stamov, D., Jessel, H., Aharoni, L., Haschke, H., Giron, J., Unger, R., Segal, E., Abu-Horowitz, A. and Bachelet, I. (2020) Visualizing the structure and motion of the long noncoding RNA HOTAIR. *Rna*, **26**, 629-636.
50. Chillón, I. and Marcia, M. (2020) The molecular structure of long non-coding RNAs: emerging patterns and functional implications. *Critical Reviews in Biochemistry and Molecular Biology*, **55**, 662-690.
51. Svergun, D.I. (1999) Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing. *Biophysical journal*, **76**, 2879-2886.
52. Svergun, D.I. and Nierhaus, K.H. (2000) A Map of Protein-rRNA Distribution in the 70 SEscherichia coli Ribosome*. *Journal of Biological Chemistry*, **275**, 14432-14439.
53. Wang, S., Chan, K.W.K., Tan, M.J.A., Flory, C., Luo, D., Lescar, J., Forwood, J.K. and Vasudevan, S.G. (2022) A conserved arginine in NS5 binds genomic 3' stem-loop RNA for primer-independent initiation of flavivirus RNA replication. *Rna*, **28**, 177-193.
54. Lee, E., Bujalowski, P.J., Teramoto, T., Gottipati, K., Scott, S.D., Padmanabhan, R. and Choi, K.H. (2021) Structures of flavivirus RNA promoters suggest two binding modes with NS5 polymerase. *Nat Commun*, **12**, 2530.

55. Fajardo, T., Jr., Sanford, T.J., Mears, H.V., Jasper, A., Storrie, S., Mansur, D.S. and Sweeney, T.R. (2020) The flavivirus polymerase NS5 regulates translation of viral genomic RNA. *Nucleic acids research*, **48**, 5081-5093.
56. Mann, M., Wright, P.R. and Backofen, R. (2017) IntaRNA 2.0: enhanced and customizable prediction of RNA–RNA interactions. *Nucleic acids research*, **45**, W435-W439.
57. Busch, A., Richter, A.S. and Backofen, R. (2008) IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics*, **24**, 2849-2856.
58. Mückstein, U., Tafer, H., Hackermüller, J., Bernhart, S.H., Stadler, P.F. and Hofacker, I.L. (2006) Thermodynamics of RNA–RNA binding. *Bioinformatics*, **22**, 1177-1182.
59. Hofacker, I.L., Fontana, W., Stadler, P.F., Bonhoeffer, L.S., Tacker, M. and Schuster, P. (1994) Fast folding and comparison of RNA secondary structures. *Monatshefte für Chemie / Chemical Monthly*, **125**, 167-188.
60. Bernhart, S.H., Tafer, H., Mückstein, U., Flamm, C., Stadler, P.F. and Hofacker, I.L. (2006) Partition function and base pairing probabilities of RNA heterodimers. *Algorithms for Molecular Biology*, **1**, 3.
61. Bernhart, S.H., Hofacker, I.L., Will, S., Gruber, A.R. and Stadler, P.F. (2008) RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics*, **9**, 474.
62. Lorenz, R., Bernhart, S.H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F. and Hofacker, I.L. (2011) ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, **6**, 26.
63. Will, S., Reiche, K., Hofacker, I.L., Stadler, P.F. and Backofen, R. (2007) Inferring Noncoding RNA Families and Classes by Means of Genome-Scale Structure-Based Clustering. *PLOS Computational Biology*, **3**, e65.
64. Pam Wang, R.A., Michelle Chen, Kristine Legaspi. (2020), Wyatt Technologies, pp. 1-4.
65. Demeler, B. and Gorbet, G.E. (2016) In Uchiyama, S., Arisaka, F., Stafford, W. F. and Laue, T. (eds.), *Analytical Ultracentrifugation: Instrumentation, Software, and Applications*. Springer Japan, Tokyo, pp. 119-143.
66. Demeler, B. and van Holde, K.E. (2004) Sedimentation velocity analysis of highly heterogeneous systems. *Analytical Biochemistry*, **335**, 279-288.
67. Meier, M., Moya-Torres, A., Krahn, N.J., McDougall, M.D., Orriss, G.L., McRae, E.K.S., Booy, E.P., McEleney, K., Patel, T.R., McKenna, S.A. *et al.* (2018) Structure and hydrodynamics of a DNA G-quadruplex with a cytosine bulge. *Nucleic acids research*, **46**, 5319-5331.
68. Panjkovich, A. and Svergun, D.I. (2017) CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. *Bioinformatics*, **34**, 1944-1946.
69. Manalastas-Cantos, K., Konarev, P., Hajizadeh, N., Kikhney, A., Petoukhov, M., Molodenskiy, D., Panjkovich, A., Mertens, H., Gruzinov, A., Borges, C. *et al.* (2021) ATSAS 3.0: Expanded functionality and new tools for small-angle scattering data analysis. *Journal of Applied Crystallography*, **54**, 343-355.
70. Putnam, C.D. (2016) Guinier peak analysis for visual and automated inspection of small-angle X-ray scattering data. *J Appl Crystallogr*, **49**, 1412-1419.
71. Burke, J.E. and Butcher, S.E. (2012) Nucleic acid structure characterization by small angle X-ray scattering (SAXS). *Curr Protoc Nucleic Acid Chem*, **Chapter 7**, Unit7.18.
72. Semenyuk, A.V. and Svergun, D.I. (1991) GNOM– a program package for small-angle scattering data processing. *Journal of Applied Crystallography*, **24**, 537-540.
73. Svergun, D.I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography*, **25**, 495-503.
74. Volkov, V.V. and Svergun, D.I. (2003) Uniqueness of ab initio shape determination in small-angle scattering. *Journal of Applied Crystallography*, **36**, 860-864.

Chapter 10. Conclusions and Future Directions

10.1 Conclusions

Throughout this thesis, I have described numerous biophysical techniques in detail: 1) Multi angle light scattering is a versatile technique that utilizes the light scattered from biomolecules in solution to determine the molar mass accurately and is often coupled to size exclusion chromatography. 2) Microscale thermophoresis, a unique technique that utilizes the measurement of a fluorescent signal corresponding to biomolecular diffusion in response to a heat source. 3) Small angle X-ray scattering, a powerful structural biophysical technique utilizing light scattering at low angles to generate a low-resolution electron density envelope corresponding to the solution structure of a biomolecule. 4) Analytical ultracentrifugation, a technique that utilizes the fundamental relationship between centrifugal force, size, shape, and density to separate and characterize biomolecules. These techniques have been fundamental to investigating non-coding viral RNA interactions and characterization throughout this thesis.

I utilized analytical ultracentrifugation, small angle x-ray scattering, and computational modeling to help elucidate the atomic structure of both MVEV and PowV terminal region non-coding RNA (Chapter 6, Nanoscale Structure Determination of Murray Valley Encephalitis and Powassan Virus Non-coding RNAs). Few studies investigate the tertiary structural organization of flavivirus genomes, despite being critical to the viral lifecycle and highly conserved within the viral family. By gaining high-resolution structures of flaviviral terminal regions, we can establish a picture of the conservation of both sequence and tertiary structure, a fundamental step towards understanding these viruses and their replication. These structures could be compared to high-resolution experimental structures in the future, to ascertain the accuracy and reliability of prediction tools.

Next, we investigated the interaction between DDX17 (Chapter 7, Human DDX17 Unwinds Rift Valley Fever Virus Non-Coding RNAs), a human RNA DEAD-box helicase, and Rift Valley Fever virus. We demonstrate that our recombinantly purified $DDX17_{135\text{-}555}$ directly binds

to both the IGR and the NCR regions of the S-segment of RVFV and unwinds both RNA in the presence of ATP. DDX17 has been implicated with numerous ncRNA interactions, and these results suggest that DDX17 could play a role in viral replication and should be further investigated in additional viral systems.

We investigated another human RNA DEAD-Box helicase, DDX3X. Previously, DDX3X was shown to interact directly with the 5′ terminal region of Japanese Encephalitis virus (Chapter 8, Human DDX3X Unwinds Japanese Encephalitis and Zika Viral 5′ Terminal Regions). Since flavivirus terminal regions are theorized to be structurally similar (not necessarily sequence similar), we investigated whether Zika virus 5′ TR could also interact directly with DDX3X. We found that JEV interacted with low μM affinity (1.66 μM), with Zika interacting with a higher affinity (7.05 μM), and that DDX3X could unwind each ncRNA in the presence of ATP. In the future DDX3X could be investigated as to its role in Zika virus infection, ideally in a cellular context.

Finally, we present a multidisciplinary approach to characterizing RNA-RNA interacting systems using the interaction between JEV 5′ TR and 3′ TR (Chapter 9, Investigating RNA-RNA interactions through computational and biophysical analysis), a phenomenon where the viral genome cyclizes to facilitate replication as a model system. We predicted the consensus secondary structure of the 5′/3′ interaction for 20 different flaviviruses, showing that the cyclization sequence is the primary interaction site. Next, we showed, for the first time, that MALS and AUC can be used to characterize an RNA-RNA interaction and quantified the binding affinity of the interaction with MST (60 nM) and further validated by thermodynamic computational analysis. This affinity was significantly weakened upon both removal and mutation of the conserved cyclization sequence. Additionally, we present the low-resolution 3D structure of both terminal regions and their complex using SAXS. This work can be used in the future to help gain insights into the binding interactions of the NS5 RdRp in three-

dimensional space to elucidate the specific mechanism by which cyclization facilitates viral replication.

In conclusion, this work uses complementary biophysical techniques to present multiple studies that characterize ncRNA in numerous viral systems, MVEV, PowV, JEV, Zika, and RVFV. Additionally, I present a pipeline for the thorough and multidisciplinary characterization of RNA-RNA interacting systems. This work helps to understand the complicated and critical interplay between viral ncRNA regions and viral replication.

10.2 Future directions and final thoughts

The most apparent future direction, but also the most difficult, is the transference of the characterization of these viral interactions from an *in vitro* system to an *in vivo* system. This transfer can be especially difficult for viral systems like JEV, which require a biosafety level 3 containment laboratory. An acceptable workaround to this hurdle is to work with a replicon system (1,2), a viral genome deficient in infectivity but replication-competent. Since we showed that Zika 5'TR interacted with DDX3X, an siRNA knockdown of DDX3X could be a feasible initial experiment to see if that interaction has any biological significance.

As for JEV (and other flaviviruses), the 5'/3' cyclization is still a relatively ambiguous process critical to genomic replication. There is evidence that this interaction can be chaperoned by the core (capsid) protein in DENV in vitro studies (3) and that the core DENV protein increases the rate by which the 5'/3' cyclization occurs (4). This chaperone activity would be worthwhile to investigate in JEV because we have already determined the affinity of the interaction to be relatively tight (60 nM), but all of our assays were incubated for ~2h to ensure the reaction was complete. Perhaps the core protein changes not only the kinetics of the cyclization reaction, but also the thermodynamics. Furthermore, our work provides evidence that while other sequences on the viral genome, namely the DAR (downstream AUG region) and UAR (upstream AUG region) can contribute to cyclization, the conserved

cyclization sequence is the primary driver *in vitro*. These minor sequences should be further investigated to elucidate their contribution to cyclization, because their proximity to the cyclization sequence could be additional sites for inhibition to combat viral replication.

Flaviviral cyclization is critical to the replication of the virus, but as mentioned earlier still not entirely understood. The primary function is theorized to be the proper positioning of the NS5 polymerase, which produces the (-) sense genome. Additionally, multiple papers assert that the NS5 polymerase is "transferred" from the 5' SLA to the 3' TR to begin transcription (5-7); however, the evidence provided is underwhelming. There needs to be more investigation into whether the NS5 polymerase is "transferred" to the 3' TR or stays bound to the 5' SLA while simultaneously interacting and transcribing the viral genome, as suggested by others (8,9). Binding to the 5' SLA could be a quality control "checkpoint" ensuring that each replicated viral genome produced by NS5 contains both the SLA at the 5' TR and the 3'TR (9). This version of the replication mechanism supports the evidence that flaviviruses have high fidelity copy of their genome, not in a sequence sense (mutations are common) but of the length of their genome (10).

While these examples are specific to viral ncRNA regions, significant questions remain unanswered regarding ncRNA structure, especially lncRNA structure. The pursuit of RNA atomic structures has made enormous leaps in the past decade with improvements to cryo-electron microscopy (11). However, as of February 23rd, 2023, there are a total of 25 RNA-only cryo-EM structures which have been deposited to databases at a resolution < 10 Å (11), and of those, only four structures exist in a resolution < 4 Å (12-15), and of those four structures at sub 4 Å resolution, only a single one is > 400 nt (15). The lack of RNA structures is evidence of a need for improved biophysical techniques for the characterization of RNA, but I do not believe that alone will propel the field forward. There must be a shift in how we discuss and search for RNA structure. Of those 25 RNA structures with a resolution < 10 Å, there is a direct correlation between the number of particles in the reconstruction, and the

resolution, more particles, lower resolution (11). This highlights an important caveat that must be considered when discussing ncRNA structure; ncRNA are flexible and can adopt different conformations in solution. If only, for example, 50,000 particles were sufficiently similar enough to merge, yet proper biophysical quality controls were performed, then the rest of the particles (discounting aggregation, of course) should still be considered relevant. Therefore, while seeking out singular atomic structures for ncRNA is worthwhile, we should not discount that ncRNA, especially lncRNA, might not exist in a biological context in a particular conformation. RNAs should be considered a conformational ensemble with potentially a few primary conformations but likely also existing in transitional states. This relates to the fundamental importance of studying RNA with multiple biophysical techniques because the more available techniques, the more evidence we can generate to help us understand the unique and complicated biological molecule ribonucleic acid.

10.3 References

1. Zhang, Q.Y., Li, X.D., Liu, S.Q., Deng, C.L., Zhang, B. and Ye, H.Q. (2017) Development of a stable Japanese encephalitis virus replicon cell line for antiviral screening. *Arch Virol*, **162**, 3417-3423.
2. Xie, X., Zou, J., Shan, C., Yang, Y., Kum, D.B., Dallmeier, K., Neyts, J. and Shi, P.Y. (2016) Zika Virus Replicons for Drug Discovery. *EBioMedicine*, **12**, 156-160.
3. Ramos-Lorente, S., Romero-López, C. and Berzal-Herranz, A. (2021) Information Encoded by the Flavivirus Genomes beyond the Nucleotide Sequence. *Int J Mol Sci*, **22**, 3738.
4. Ivanyi-Nagy, R. and Darlix, J.L. (2012) Reprint of: Core protein-mediated 5'-3' annealing of the West Nile virus genomic RNA in vitro. *Virus Res*, **169**, 448-457.
5. Fajardo Jr, T., Sanford, T.J., Mears, H.V., Jasper, A., Storrie, S., Mansur, D.S. and Sweeney, T.R. (2020) The flavivirus polymerase NS5 regulates translation of viral genomic RNA. *Nucleic acids research*, **48**, 5081-5093.
6. Filomatori, C.V., Iglesias, N.G., Villordo, S.M., Alvarez, D.E. and Gamarnik, A.V. (2011) RNA sequences and structures required for the recruitment and activity of the dengue virus polymerase. *Journal of Biological Chemistry*, **286**, 6929-6939.
7. Filomatori, C.V., Lodeiro, M.F., Alvarez, D.E., Samsa, M.M., Pietrasanta, L. and Gamarnik, A.V. (2006) A 5′ RNA element promotes dengue virus RNA synthesis on a circular genome. *Genes & development*, **20**, 2238-2249.
8. Choi, K.H. (2021) The Role of the Stem-Loop A RNA Promoter in Flavivirus Replication. *Viruses*, **13**, 1107.
9. Lee, E., Bujalowski, P.J., Teramoto, T., Gottipati, K., Scott, S.D., Padmanabhan, R. and Choi, K.H. (2021) Structures of flavivirus RNA promoters suggest two binding modes with NS5 polymerase. *Nature Communications*, **12**, 2530.
10. Selisko, B., Papageorgiou, N., Ferron, F. and Canard, B. (2018) Structural and Functional Basis of the Fidelity of Nucleotide Selection by Flavivirus RNA-Dependent RNA Polymerases. *Viruses*, **10**.
11. Ma, H., Jia, X., Zhang, K. and Su, Z. (2022) Cryo-EM advances in RNA structure determination. *Signal Transduction and Targeted Therapy*, **7**, 58.
12. Zhang, K., Li, S., Kappel, K., Pintilie, G., Su, Z., Mou, T.-C., Schmid, M.F., Das, R. and Chiu, W. (2019) Cryo-EM structure of a 40 kDa SAM-IV riboswitch RNA at 3.7 Å resolution. *Nature communications*, **10**, 5511.
13. Bonilla, S.L., Sherlock, M.E., MacFadden, A. and Kieft, J.S. (2021) A viral RNA hijacks host machinery using dynamic conformational changes of a tRNA-like structure. *Science*, **374**, 955-960.
14. Weinberg, Z., Barrick, J.E., Yao, Z., Roth, A., Kim, J.N., Gore, J., Wang, J.X., Lee, E.R., Block, K.F. and Sudarsan, N. (2007) Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic acids research*, **35**, 4809-4819.
15. Su, Z., Zhang, K., Kappel, K., Li, S., Palo, M.Z., Pintilie, G.D., Rangan, R., Luo, B., Wei, Y. and Das, R. (2021) Cryo-EM structures of full-length Tetrahymena ribozyme at 3.1 Å resolution. *Nature*, **596**, 603-607.