12-10-2018

Structure-Function Relationships of the Nucleocapsid in Vesicular Stomatitis Virus

Ryan Gumpper

STRUCTURE-FUNCTION RELATIONSHIPS OF THE NUCLEOCAPSID IN VESICULAR STOMATITIS VIRUS

by

RYAN GUMPPER

Under the Direction of Ming Luo, PhD

ABSTRACT

Understanding the basic structure-function relationships that guide the biology of various pathogenic viruses is the cornerstone in the development of antiviral therapies. One group of highly pathogenic viruses, which desperately needs novel antiviral therapies, are Negative Strand RNA viruses. This group of viruses contains some of the most highly pathogenic and pandemic viruses known: like Ebola, Influenza, Marburg, and Rabies viruses. Between 2013 and 2016, during the latest endemic outbreak of Ebola, the World Health Organization estimated that it caused over 11,000 deaths. However, without in-depth understanding of the basic science that guides the replication of these viruses, design of targeted therapies is hard. Utilizing the model virus for Negative Strand RNA viruses, Vesicular Stomatitis virus, this work studies the

structure-function relationships of the nucleocapsid protein. Since the nucleocapsid encapsulates the virus' genomic material this protein is increasingly important study. Furthermore, the virus cannot undergo replication if the polymerase does not recognize the nucleocapsid for viral RNA synthesis. This work takes a three-pronged approach when investigating the nucleocapsid: development of a novel antiviral to target the nucleocapsid, the use of structural information to study the possible interactions between the nucleocapsid and the polymerase, and a study of the genomic constraints placed on polymerase activity by the nucleocapsid. Utilizing an already characterized class of compounds, polyamides, we showed that it is possible to target the genomic RNA in the nucleocapsid of a Negative Strand RNA virus and inhibit viral replication. This is the first known instance of this occurring in Negative Strand RNA viruses. To examine the structural interactions between the nucleocapsid and the polymerase a flexible loop was hypothesized to be integral toward the formation of an active polymerase complex. Alanine scanning deletions yielded an abolishment in polymerase activity, which could be restored after compensatory mutations were allowed to occur. Finally, the nucleotide content sequestered within the nucleocapsid has a direct effect on the processivity of the polymerase complex, and ushers in a paradigm changing model for how one examines Negative Strand RNA virus activity and the evolutionary constraints placed on the genomic RNA. Merging the findings of these studies together, understanding the structure-function relationships of the nucleocapsid can yield a lot of knowledge about the viral life-cycle of the entire Negative Stranded RNA virus family.

INDEX WORDS: Negative Strand RNA virus, Nucleocapsid, Antiviral therapies, Vesicular Stomatitis Virus

STRUCTURE-FUNCTION RELATIONSHIPS OF THE NUCLEOCAPSID IN VESICULAR STOMATITIS VIRUS

by

RYAN GUMPPER

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2018

STRUCTURE-FUNCTION RELATIONSHIPS OF THE NUCLEOCAPSID IN VESICULAR STOMATITIS VIRUS

by

RYAN GUMPPER

Committee Chair: Ming Luo

Committee Members: Jenny Yang

Gregory Poon

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

November 2018

DEDICATION

I would like to dedicate this dissertation to my Mom, Dad, Brother, and fiancé, Suela. Without the love and support of all of you, none of this would be possible.

ACKNOWLEDGEMENTS

First, I would like to acknowledge my advisor, Dr. Ming Luo, for always leading by example and teaching me what it means to be a true scientist. I would like to thank him especially for taking me into his lab when I thought my entire PhD was in jeopardy and giving me the chance to succeed. Furthermore, I would also like to thank him for the ongoing moral and academic support through this process and allowing me the space to explore new ideas not only scientifically, but also personally. Without his help, none of this would have been possible.

I would also like to thank both of my committee members Dr. Gregory Poon and Dr. Jenny Yang. To Dr. Yang whom I have interacted with since my first day in graduate school in her Protein Structure and Function class. From being on my oral exam committee, to putting together the wonderful joint group meetings, and who has offered nothing but amazing scientific and professional advice all the way through my graduate career. Also, to Dr. Gregory Poon who has offered a plethora of moral support and has offered many insights into the ins-and-outs behind science and the scientific process. I always know that I can come to you for your sage advice.

I would also like to thank Dr. Jiafeng Geng and Dr. Kednerlin Dornevil, both members of my former lab. Not only did both of you teach me all of the necessary laboratory techniques but were also an invaluable part of my growth as a scientist when I first started graduate school. As a freshmen graduate student, having senior members take you under their wing, like both of you did, did more to shape my career than you can imagine. Also, to the current members of my lab, especially Weike (Eric) Li and Ross Terrell. To all the inspiring conversations, both scientific

and non-scientific while toiling away for the endless hours in lab. Both of your friendships mean the world to me and helped keep me sane when times were tough.

I would also like to thank my family: my Mom, Dad, and Brother. Being far apart these past 5 years has been tough, but without your constant love and support throughout this entire process I surely wouldn't have been able to do it. All of you have raised me to be an independent thinker and a hard worker, and without these traits which have been so ingrained into the fabric of my being I would not be where I am today. Thank you for traveling all those times to see me, as graduate school gets very busy at times, and I am eternally grateful for all your sacrifices in time, money, and heartache at being far away. Sometimes achieving something worth doing takes a lot of sacrifice but know that my love for you all is unconditional, and I look forward to the days we can spend more time together. Without you guys as my foundation, I wouldn't have been able to build the tower that you see today. I love you guys and look forward to the future!

Finally, I would like to especially thank my partner in crime, the love of my life, and other-half, my fiancé Suela. To this day, your love and devotion continues to inspire me to reach to new heights. Without you by my side every step of the way life would not be as happy, fulfilling, or fun. Growing up together through graduate school will be a time in our lives that we will always cherish. It was during this time that our relationship blossomed, and our love grew to enormities that I could never comprehend to exist. Through all the failed experiments, the late nights and early mornings in the lab, and plenty of tears you have been my rock. Every day with you is filled with adventure and love, and without your emotional and professional support I would not be half the man that I am today. This journey has not been easy, and the road has not been straight, but I cannot think of any other person I would rather have by my side than you.

Thank you for being you and I will continue to love you as this next exciting step in our lives begins. I love you, and I always will—here's to the next adventure.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Reverse Transcription qualitative Polymerase Chain Reaction (RT-qPCR)
Ribonucleic Acid (RNA)
Deoxyribonucleic Acid (DNA)
Human Immunodeficiency Virus (HIV)
messenger Ribonucleic Acid (mRNA)
Center for Disease Control (CDC)
Acquired Immunodeficiency Syndrome (AIDS)
Negative Strand RNA Viruses (NSVs)
Respiratory Syncytial Virus (RSV)
Complementary Ribonucleic Acid (cRNA)
Viral Ribonucleic Acid (vRNA)
Ribonucleoprotein (RNP)
Transfer Ribonucleic Acid (tRNA)
Cytosine paired with Guanine (CpG)
Toll-like Receptors (TLRs)
International Committee on Taxonomy of Viruses (ICTV)
Protein Databank (PDB)
National Institutes of Health (NIH)
Human Papilloma Virus (HPV)
Negative Strand RNA Virus (NSV)
Codon Usage Bias (CUB)
Codon Pair Bias (CPB)
Viral RNA Dependent RNA polymerase (vRdRp)
Vesicular Stomatitis Virus (VSV)
Global Health Observatory (GHO)
World Health Organization (WHO)
Federal Drug Agency (FDA)
Nucleocapsid-like Particle (NLP)
Vesicular Stomatitis Virus (VSV)
Nucleoprotein (N)
Phosphoprotein (P)
Matrix-protein (M)
Glycoprotein (G)
Large protein (L)
RNA-dependent RNA polymerase (RdRp)
Tyrosine (Tyr)
Cryo-Electron Microscopy (Cryo-EM)
Double Stranded DNA (dsDNA)
Single Stranded RNA (ssRNA)
Single Stranded DNA (ssDNA)
Double Stranded RNA (dsRNA)
Imidazole (Im)
Pyrrole (Py)
Thermal Shift Assay (TSA)

1 GENERAL INTRODUCTION: WHY STUDY VIRUSES?

In biology, viruses represent an intellectual duality: they are so incredibly simple, but also so frustratingly mysterious. They have often perplexed Nobel prize winning researchers, while only containing the basic biological building blocks to replicate. These parasitic molecular machines are so vast in their diversity and display a dizzying array of functionality by infecting every living thing on the planet (1). Meanwhile, they simultaneously lack basic biological functions like a metabolism and ability to replicate outside of a host. It is through the study of this "contagium vivum fluidum" (contagious living fluid) (1, 2) that a relatively new scientific field has emerged, Virology. Virology has made huge strides in understanding and controlling pandemic pathogens, but at other times has completely reinvented scientific dogma. It is in the spirit of exploration and human ingenuity that makes this field of study so patently interesting to study.

1.1 General History of Virology

From their initial discovery in 1892 by Russian scientist Dmitri Ivanofsky until modern day, viruses represent a biological enigma (1, 3). In the late 19th century, the prevailing theory of infection was already well established by the works of Koch, Lister, and Pasteur. Koch's postulates clearly outline a model for microscopic bacterial infection (1):

(a) The organism must be regularly found on the lesions of the disease.

(b) The organism must be isolated in pure culture.

(c) Inoculation of the pure organism into the host should initiate the disease.

(d) The organism must then be recovered once again from the lesions of the host.

At this time, this would have been considered scientific dogma. However, things changed with the invention of the Chamberland filter. Invented by Charles Chamberland, a colleague of

Pasteur, the Chamberland filter was a revolutionary porcelain filter with pores small enough to filter out bacterial agents. It was during this time that Dmitri Ivanofsky was working on a disease that was ravaging the Russian tobacco crop (2, 3). He isolated the sap from the infected plant, which would normally be able to infect another plant, and passed it through a Chamberland filter. He hypothesized that the small pores in the filter would remove the infectious agent. However, this was not the case. The filtered material could still infect a healthy plant. Ivanofsky reported this information to the National Academy of Sciences, but the scientific understanding of the time did not allow him to make any overly zealous claims.

It was not until a couple of years later that Martinus Beijerink unknowingly and independently verified the work of Ivanofsky (in 1898) by repeating the same experiment (4). However, Beijerink carried Ivanofsky's work one step further: he showed that a serial dilution of the filtered sap material, could regain its potency after replication in the host. He later found that the agent could not be isolated and cultured, violating Koch's principles, but needed the living cells or tissue to be able to reproduce (4). This kicked off a period of scientific debate into the nature of this infectious liquid. Was it a toxin? An enzyme? Or a particle too small to be seen by a light microscope? It was during this time, that the literature started to identify numerous accounts of these "filterable infectious agents" which violated Koch's postulates and produced an amalgam of diseases (5-9). They started to call them viruses, taken from Latin meaning "slimy liquid or poison"(9).

The identity of the filterable infectious liquid, was finally settled with the invention of the electron microscope in 1937 when the first electron micrographs were taken of the communicable sap and small rod-like particles were identified. This was to become the first image of a virus, in this case the identification of Tobacco Mosaic Virus (6). It was this ground-

breaking technique, along with the development of the plaque assay in 1917 (5), which kicked off modern Virology as we know it today.

Since their discovery, viruses have represented significant scientific paradoxes being both alive and not alive, being eloquently simple and frustratingly complex, and capable of massive death on the microbiological and macrobiological scale. It is only through the understanding of the basic biological driving forces behind these primitive parasites that scientists can begin to unwind these oxymora and usher in a new age of scientific enlightenment and human health. With the constant threat of emergent pandemic diseases threatening human civilization, it is of vital importance to thoroughly study and understand these perplexing biological materials. As Sun Tzu said in the Art of War, "If you know the enemy and know yourself, you need not fear the result of a hundred battles. If you know yourself but not the enemy, for every victory gained you will also suffer a defeat. If you know neither the enemy nor yourself, you will succumb in every battle." This is why we study viruses.

1.2 Viral Evolution, Classification System, and Antiviral Therapies

To fully appreciate the scope of this dissertation, it is important to have a competent understanding of some basic principles of virology. These basic principles, serve as a signpost to constantly compare and validate our exploration of the cutting edge basic science in the context of the established scientific principles. The principles of viral evolution, classification, and discovery of antiviral therapies are directly applied throughout this dissertation. This section explores these basic principles to establish an understanding and common linguistic framework as applied to the research in this dissertation.

1.2.1 Viral Evolution

Viruses represent a biological conundrum. This mystery is exacerbated when examining the evolution of viruses. Since viruses lack some of the basic machinery to be called a "living" organism, there are several evolutionary theories which pervade the scientific literature. Were viruses the first form of life, arising from RNA and eventually becoming more complex? Or did viruses co-evolve with the evolution of cells and bacteria, enacting out their parasitic existence from the start? Or perhaps they are a byproduct of ancient genetic material that escaped the cell, still needing it for replication? These are all valid hypotheses and are actively being studied and discussed. If one can understand viral evolution, then perhaps this knowledge can be used to predict novel pathogenic viruses arising from extraneous sources. Furthermore, with the vast amount of genomic data currently available, these predictive abilities will only continue to get better. Very much like evolution is at the heart of biology, the application of evolutionary principles is also at the heart of virology. It is the guiding force which informs all biological function.

The first theory of interest is the "virus first" hypothesis. As this hypothesis suggests, viruses, particularly RNA viruses, predated cellular existence (10-12). This theory postulates that there was a pre-cellular world which belonged to a batch of ancient viruses. It was these ancient viruses that evolved and developed the molecular tools needed for cells to function. This theory is supported by the fact that viral genomes and proteins are often vastly different from their cellular counterparts. These sequence specific viral genes, do not share homologous proteins in other living materials. This hypothesis infers that before cells arose, there was a primordial genetically diverse pool in which freely intermixing components were assembled. This primordial pool eventually gave rise to modern cells and modern viruses, but the essential

building blocks arose from here (12). While this hypothesis states that the modern parasitic activity later evolved once cells took hold, a large gap in the evolutionary understanding remains. Also, all viruses need a cell to replicate. While it does not disprove of this hypothesis, the common ancestor between cells and viruses could be the same, due to this reason this hypothesis has been succinctly challenged and further critiqued (13).

Another way that viruses could have arose is from a co-evolution with cells. This is called the "reduction hypothesis". Much like the "virus first hypothesis", this premise suggests that there was a pre-cellular world and viruses co-evolved alongside cells (10, 14). This co-evolution means that viruses are just reduced parasites, ranging from very simple, like RNA viruses, to vastly complex, like some double stranded DNA viruses. As cells got more advanced, the viruses were able to evolve along with cells becoming more advanced themselves. This hypothesis has recently garnered more attention with the discovery of the very large mimiviruses, megaviruses, and most recently the Pandoravirus (15-17). These viruses are giant with the genome of Pandoravirus containing around 2.5 million bases and encoding for 2,500 genes. The viruses share many features with other parasitic bacteria, and have added clout to this hypothesis of viral evolution (17). Furthermore, with the discovery of these viruses it has sparked the discussion to add viruses as the fourth domain of life (18). Due to their enormous complexity it could be hard to see how these viruses, are not something that should be seriously considered as a life-form. However, the presence of these viruses can also lend credence to the next hypothesis of viral evolution.

The third hypothesis of viral evolution is called the "escape hypothesis" (11, 13). This hypothesis proposes that viruses are a part of the host cell. The genetic material surrounding viral functions, somehow escapes from the host cell. However, for replication of the genetic material

to take place, it then needs to return to the host cell. These “viruses”, or “viroids”, then can further evolve gain of function through horizontal gene transfer. Horizontal gene transfer occurs when one organism, in this case a virus, steals a gene from the host cell which then becomes ingrained into the viral replication cycle (19, 20). While this is most certainly has occurred, it again runs into a similar problem as the “virus first” hypothesis: How does one explain that the vast majority of viruses are non-homologous to cellular organisms (10, 13, 21)?

While a version of every hypothesis carries a kernel of truth, the origins of viruses and their evolutionary path most-likely lies somewhere between all these hypotheses. One interesting concept, which would link all life together and viral evolution, is the concept of the virocell (10, 12, 22, 23). The virocell is a cell that has been infected by a virus and its whole purpose is to produce virions. In the context of a cell, the virus is “alive” and replication starts to occur, effectively turning the infected cell into an entire new type of organism. This concept places the evolution of viruses just after the evolution of the ribosome and proteins, but before the last common ancestor of the modern cell. Viewing viruses from this lens is interesting because viruses are then “alive” during the intracellular phase of the life cycle (23). While this is still surely up for debate, viewing the virus in the context of a ribosome, is an interesting intellectual take on viral evolution.

While viral evolution is a widely debated topic, studying viral evolution is important to understand both the mechanisms of viral replication and how these viruses will continue to change with time. Additionally, in investigating the origins of life, viruses have played a large role in unlocking many mysteries surrounding seemingly intractable problems. Until now, the diversity of the viral population has not been discussed, and understanding the basic

classification system, which is an evolutionarily driven discussion, is the topic of the next section.

1.2.2 Viral Classification System

In trying to better understand viruses, scientists have come up with ways to classify them. Being so diverse, with a wide range of hosts, symmetries, and methods of replication, viral classification has proved to be an interesting problem for a long time. Initially, virologists were classifying viruses by the nature of the disease and the host that the virus infected (24). However, this turned out to be a problem for a number of reasons: a plethora of viruses could cause similar symptoms to appear throughout the course of a disease, making classification extremely difficult; the structure of different virions can be radically different, but still infect the same host while possibly causing similar symptoms; and the molecular material which the virus uses to replicate can drastically vary among viruses that infect the same host (25). For the first 60 or so years of virology, the nature of classifying viruses was hotly debated. Should it even be done, as viruses are intracellular obligate parasites, and are not in the traditional sense alive? However, as more viruses were discovered the need for a classification system become inherently relevant (25).

In 1971 the seminal paper on the widely used classification system, for which he won the Nobel Prize in 1975, was published by David Baltimore (26). In the paper, he proposed a classification system based on the method of replication by viruses. Once viruses were classified by the method of replication, then further classification could take place based on the structure of the virion and host preferences. Baltimore first suggested 6 different families to classify viruses (**Figure 1.1**). This was later amended to include a 7th class as new data arrived, and a new family of viruses was needed (27, 28). Utilizing the genomic structure and method of viral replication was genius, given that DNA sequencing was very much still in its infancy and very

little was understood about viral life cycles at the time. Even more impressive, was that this classification system still holds true today, and has largely been able to characterize all the new viruses discovered. Remember, retroviruses like HIV would not be discovered for another 10 years, but Baltimore was able to understand the literature so thoroughly and propose this class of viruses.

The Baltimore classification system currently contains 7 groups of viruses all based on the genetic material and method used for viral replication. Each group can then be split and classified further based on the structural properties of the genome (i.e. linear, circular, or segmented genomes) and structural properties of the virion (1). Group I viruses (**Figure 1.1**), use a double stranded DNA genome to replicate and proceed to protein synthesis using an mRNA intermediate. Group I viruses have a varied genomic structure, containing linear, circular, segmented and circular, and segmented and linear genomes. This group contains a wide variety of pathogens, infect seemingly all types of living things (even some other viruses), and is classified based on morphology instead of sequence similarity. It currently contains 3 orders, which are further divided into 31 families of virus. Some clinically relevant viruses include the Herpesviridae and Poxviridae. The family of Herpesviridae is responsible for chicken pox, while the Poxviridae family hosts viruses like small pox, both of which have been nearly eradicated with vaccines (29-31).

Group II viruses (**Figure 1.1)** contain a single positive stranded DNA genome. To make mRNA, the virus must first produce the template strand (- strand) and then can transcribe the mRNA for protein synthesis. This means that this virus goes through a double stranded DNA intermediate before proceeding to protein production. This group of viruses is most famous for bacteriophage Phi X174, which was discovered to have a single stranded DNA genome in 1962

(32). This genome was the first DNA genome to be completely sequenced by Fred Sanger in 1977 (33). This group of viruses has proven hard to classify and is home to a diverse group of pathogens found throughout the environment. The morphological and genetic differences between the taxa of these viruses are often large, but efforts have been made to better classify these orders. The genomes of these viruses are notoriously small, and are mainly circular, with a few exceptions. One family of clinically relevant single stranded DNA viruses is Parvoviridae. Parvoviridae is home to the parvovirus, in which Parvovirus B19 was found to infect children causing a rash, while canine parvovirus can be highly lethal in unvaccinated dogs and is spread through feces (34, 35).

Group III viruses (**Figure 1.1**) are double stranded RNA viruses which will transcribe RNA from the genome for both protein synthesis (mRNA) and viral replication. The genomic structure of this group of viruses is highly variable and can be either segmented or non-segmented, with the segmented viruses containing up to 12 segments. Double stranded RNA viruses can infect a wide range of hosts including plants, animals, fungi, and bacteria. Furthermore, the variability of the structures of the virion in this class of viruses shows remarkable diversity. One clinically relevant pathogen belonging to this group of viruses is the rotaviruses belonging to the family Reoviridae. Rotaviruses are responsible for severe gastrointestinal distress and diarrhea in babies and young children and in 2013 was responsible for up to 215,000 deaths in the developing world (36, 37).

Moving down the taxonomic groups, Group IV viruses (**Figure 1.1**) utilize a single (+) stranded RNA genome. Unlike many other viruses, this group can utilize the packaged (+) stranded RNA directly for protein synthesis. However, for replication to take place the negative strand needs to be synthesized, so more positive strands can be made for replication to be

complete. Largely classified based on the morphology of the RNA dependent RNA polymerase, the (+) stranded RNA viruses are home to the most RNA viruses, currently with 3 orders and 34 different families being recognized. This group of viruses is perhaps most famous for the family Picornaviridae which is home to viruses like poliovirus, rhinovirus (the common cold), and Hepatitis A virus (38).

Skipping over Group V viruses (**Figure 1.1**), as section 1.3 is entirely devoted to them, Group VI viruses also contain a single (+) stranded RNA genome. However, the mode of replication is very different from the Group IV viruses. Group VI viruses utilizes reverse transcription during its life cycle to mediate a DNA intermediate. This DNA intermediate is then utilized to create new (+) stranded RNA genomes. In the case of the widely known family of Retroviridae the DNA intermediate is incorporated into the host genome, known as a provirus, before replication takes place. The most famous virus in the group of viruses is HIV, which according to the Centers for Disease Control effected 36.7 million people and an estimated 1 million people died from AIDS related illnesses in 2016 alone (39). While retroviruses have certainly caused much sorrow and death, the discovery and utilization of the reverse transcriptase enzyme has yielded many ground-breaking discoveries in biomedical research and is actively used throughout the research community today (40, 41).

The last group of viruses, Group VII viruses (**Figure 1.1**), also utilize a reverse transcriptase and are another group of double stranded DNA viruses. However, unlike Group I viruses, these viruses also use a reverse transcriptase for replication. Group VII viruses will not only transcribe mRNA for protein synthesis, but also transcribe pre-genomic RNA. This pre-genomic RNA, through reverse transcription, acts as a template and is transcribed back to DNA to replicate the genome. This is a relatively small group of viruses with two different families.

The most famous virus of this group is Hepatitis B virus, which is responsible for severe liver infections and can lead to hepatic cirrhosis and possibly hepatic cancers (42). While there is a vaccine available for Hepatitis B virus, if some individual contracts the virus there is no known cure.

Viral classification is important to understand the different modes which viruses employ to replicate within a host. Virus taxonomy is still an ever changing topic and is continually monitored and changed by the International Committee on Taxonomy of Viruses (ICTV) (43). The ICTV has 4 main objectives as posted on their website: "to develop an internationally agreed taxonomy for viruses; to establish internationally agreed names of virus taxa; to communicate the decisions reached concerning the classification and nomenclature of viruses to virologists by holding meetings and publishing reports; to maintain an official index of agreed names of virus taxa" (44). As the ability for us to sequence genomes faster and cheaper continues to accelerate, the discovery of novel viruses will concomitantly quicken as well. Classification of viruses is becoming an important role, not only for monitoring viral evolution and infection, but also for adequately identifying and understanding novel viruses within the context of the known scientific literature.

1.2.3 Antiviral Therapies

The practice of understanding the structure-function relationships at the protein level has led to a giant leap in our basic understanding of the biological universe. This impact has not only been felt in the sciences but can also be directly correlated with the increase in life expectancy and increase in quality of life. From 2000 to 2015 the average world life expectancy has risen by 5 years (**Figure 1.2A**)(The Global Health Observatory (GHO) data repository under World Health Organization (WHO)) (45). Furthermore, we have also seen an explosion in the number

of structures deposited in the Protein Data Bank (PDB) (**Figure 1.2B**) (46). When looking at the correlation between deposited protein structures and global life expectancy, one notices a significant positive correlation between the two (**Figure 1.2C**). While of course this is not a causative effect, but it is evidence that as our knowledge of protein structure has amassed the ability to apply this knowledge in a medically meaningful way has also gotten better. For example, the Federal Drug Association has approved a record number of new medications in 2017 with 46 novel drugs approved (47). This implies that understanding the basic science behind these structure-function relationships is not only an intellectual curiosity, but it is of vital importance towards improving the over-all human condition.

While the understanding of structure-function relationships has continued to improve, there are certain areas where the basic science is lacking. One such area is in the development of antiviral drugs and antiviral treatments. With such pandemic outbreaks as the Ebola outbreak in 2014 (48) and the constant threat of the annual influenza outbreak, which costs U.S. workplaces approximately $7 billion annually (49), novel therapies are desperately needed. Luckily enough, in the cases outlined above, there has been a large push in recent years by the NIH to meet these needs (50). A success story in the application of structure-function relationships in drug discovery, was with the innovation of neuraminidase inhibitors (51-53). These inhibitors target neuraminidase activity in influenza, inhibiting the virus' ability to reproduce by blocking the budding step (51). Two such drugs are now approved for use in the U.S. and Europe, oseltamivir and zanamivir, to treat both Influenza A and B. However, the efficacy of this class of drugs is widely debated and with treatment being most efficacious when the patient is treated early in the infection (54-57). Another class of drugs that has seen a recent success in antiviral therapies are polyamides. Polyamides have been long used to target the

minor groove of double stranded DNA in a sequence specific manner (58-60). Polyamides are a macromolecule with a repeating set of amide bonds. The interesting thing about these compounds is the chemical moieties can be changed to modulate specificity for a specific DNA sequence (60-62). Since these compounds can be designed for a specific DNA sequence, this leaves little room for off-target effects. Recently this class of drugs has been used to target the double stranded DNA virus, Human Papilloma virus (HPV), which is widely known to be implicated to cause cervical cancer (63-65). These compounds have been widely studied in *in vitro* and are hopefully going to make the move to being used in the clinic. For the context of this work, polyamides have been applied as a novel antiviral against a negative strand RNA virus, representing a new use for this highly versatile class compounds.

One cannot mention antiviral therapies without talking about vaccines. Vaccines are the most efficacious defense against viruses from all groups. Utilizing the power of our own immune system, vaccines have been extremely successful in preventing common diseases that were widespread for most of human history like smallpox, polio, whooping cough, measles, mumps, and rubella (66-69). Perhaps the biggest success story of vaccines is that of smallpox. Since 1980, the World Health Organization declared smallpox to be eradicated (69). This was a huge step for humankind as smallpox was the culprit of many pandemics throughout human history. While these vaccines have been very important towards bettering human health and longevity, further development is needed to combat other highly pathogenic viruses. One promising case, which arose from the Ebola epidemic from 2013-2016, which reportedly killed 11,310 people with a mortality rate of 39.5%, is the invention of an Ebola vaccine (70, 71). In early vaccination trials, this vaccine is almost 100% effective at offering protection against Ebola infection (70, 71). These success stories around vaccines, are important to be shared with the

current combative climate surrounding vaccination. Since the anti-vaxxer movement has taken hold, diseases that have not been seen in decades, due to the success of vaccines, are once again on the rise (72, 73).

While these results from antiviral and vaccines are certainly encouraging, for major advances to be made, and novel antiviral therapies to be produced, further understanding of the basic structure-function relationships of these viruses are vitally important to future generations. It is the hope of this work that through the study of the basic structure-function relationships, which guide viral replication, that guiding principles can be elucidated to better understand and design novel antiviral therapies. Utilizing both biochemical and biophysical techniques to unravel the structural complexity of these molecular machines can have a significant impact on human health, longevity, and well-being.

1.3 Negative Strand RNA Viruses

Negative strand RNA viruses (NSVs), or Group V viruses, are responsible for some of the most pandemic human pathogens known. This family of viruses is home to many of the well-known and publicized hemorrhagic viruses such as Ebola, Marburg, and Lassa viruses, which can horrifyingly induce hemorrhaging from the gums, eyes, nose, and other orifices (**Figure 1.3B**). They are also famously involved in thousands of deaths (i.e. the Ebola outbreak in 2013-16) (74, 75). Another pathogen which is a part of this family is the seasonal flu (Influenza A and B), which the Centers for Disease Control (CDC) estimates is responsible for anywhere between 12,000 and 56,000 deaths in the U.S. every year (76). Furthermore, this family also includes diseases which are deemed largely under control due to the development of vaccines (i.e. Measles, Mumps, Rubella, Rabies), but are now making a return with the “antivaxxer” movement and rapidly developing resistance (73). Also, Human Respiratory Syncytial Virus

(RSV) is an NSV which is the leading cause of lower respiratory tract infections during infancy and early childhood. It will usually infect almost all children by the time they reach 2-3 years of age and is one of the top reasons for childhood hospitalization (77). As one notices, NSVs are responsible for a wide range of human illnesses, hospitalizations, and death. It is fascinating that this group causes such a wide range of symptomatic phenomena and is also extremely promiscuous in its choice of hosts: infecting vertebrates, arthropods, and even plants.

Despite being so promiscuous, this group of viruses share several common traits. As mentioned in the chapter on virus classification, viruses can be classified by their mode of replication. NSVs utilize a negative sense genome as its genetic material. Unlike positive strand RNA viruses, which can directly utilize their RNA for translation of viral proteins, NSVs must go through a transcription step in order to create the necessary mRNA for protein translation (**Figure 1.3A**) (78, 79). In doing so, instead of relying on the host's polymerase machinery, this group has evolved to bring along their own, called the viral RNA dependent RNA polymerase (vRdRp) (80-82). Additionally, for this group of viruses to replicate, it must first transcribe a sufficient amount of viral proteins, and then initiate replication by creating a complete complementary genome (cRNA) (**Figure 1.3A**). Then the polymerase complex must re-copy the cRNA back to the normal negative stranded genome (vRNA). This mode of replication is unique to Group V viruses, and they have come up with interesting ways to achieve this goal.

Another commonality of this group of viruses is that the genome during entry, assembly, transcription, and replication, must be encapsidated by a capsid protein (83-88). This means that at every point of the life-cycle, the viral genome is protected by a capsid protein. Moreover, for the polymerase to partake in transcription/replication it must bind to the capsid protein and access the encapsidated RNA. This makes accessing the genome an imperative step for

transcription and replication to occur. This is unique for this group of viruses: the vRdRp can only recognize the protein RNA template for transcription and will not transcribe the naked genomic RNA. This adds an extra layer of regulation in the production of RNA and protein levels in this group of viruses.

This family of viruses can also be split into two distinct groups based on the structure of the genome: segmented and non-segmented. Segmented NSVs, as the name implies, have a segmented genome, with each segment encoding for a specific protein or proteins. While non-segmented NSVs have a continuous linear genome with each gene being separated by intergenic regions controlling transcription. For segmented NSVs, each segment is individually encapsidated and then packaged separately into the virion, and often exhibit a circular helical structure (87). This means that each ribonuclear protein (RNP) in segmented NSVs acts as a functionally independent moiety during transcription and replication. Whereas non-segmented NSVs contain a continuous linear genome which is usually packaged into a tight helical structure (85, 89, 90). Since all of the genes are part of a continuous RNA strand, this means that each gene is sequentially transcribed allowing for a protein gradient to form (91). While this represents two different modes of transcriptional control, it is interesting that for segmented NSVs, this allows for substantial levels of recombination to occur during a superinfection. This possibly leads to the generation of species hopping strains, which often observed during influenza infections (92, 93). Whereas in non-segmented NSVs this phenomenon does not occur.

1.3.1 Evolutionary Control of Genetic Drift

One interesting phenomena, actively discussed and explored in this dissertation, is how NSVs controls their genetic drift. Due to the nature of RNA transcription, NSVs often have surprisingly high mutation rates. This has been traditionally explained by the fact these viruses

bring along their own transcriptional machinery. Having a high mutation rate can serve as an evolutionary advantage due to the presence of quasi-species and possible beneficial mutations (94, 95). It has been noted that Paramyxoviridae exhibits a high enough mutation rate to completely randomize the entire genome up to 6 times per year (96). However, the observed mutational rate is only around 6 nucleotides per year (96). This creates a dichotomy, and begs the question: how can the genomic sequence of these viruses be so stable, in the presence of such a high mutation rate?

One way that these viruses could adequately control this phenomenon, is through the optimization of their Codon Usage Bias (CUB) and Codon Pair Bias (CPB). For efficient translation by the host, the virus must match the host's CUB. CUB can be defined as the percentage of a codon used for a selected amino acid during translation. This bias has been found to differ across species and throughout phyla. This is usually attributed to the abundance of tRNAs that are available for the host to use. CPB is related to how often codons are synonymously paired with each other throughout a sequence, which can be a direct consequence of CUB. It has been widely shown that positive strand RNA viruses match their CUB/CPB with that of the host organism's (97). Furthermore, it has also been shown that one can create a live attenuated virus, just by changing the CUB/CPB of Poliovirus (a positive stranded virus), and not the encoded proteins (97, 98). This technique has been used for the past decade, and is currently under intense investigation (99). However, while this works for positive stranded RNA viruses, this does not mean it will be the same for NSVs. Not only will the host's translational machinery play a role in regulating genomic processivity, but also NSVs bring along their own transcriptional machinery for mRNA synthesis. This makes it another checkpoint for transcriptional/replication control. Furthermore, the CUB/CPB in NSVs, for some reason do not

actually match with their host's (**Table 4.4** in Chapter 4). Leading to another level of mystery as to how they regulate their genomic drift.

Another level of evolutionary constraint against NSV genomes, is the suppression of CpG. It has been widely noted, especially in DNA viruses, that CpG is widely avoided in viral and pathogenic infections. This is usually due to two different factors: avoiding genomic methylation and evasion of the innate immune response. In DNA CpG is often avoided due to the methylation of C to a T, circumventing a level of epigenetic control and or mutational hotspots. Also, CpG is avoided to help the parasite evade the innate immune response through interaction with TLR9 (100-103). While NSVs have been shown to avoid CpGs, this correlation has been shown to be not as strong—as in a positive strand RNA virus like polio (104, 105). However, while it is widely understood in the context of DNA, there is contradictory evidence in the literature as to whether this actually applies to RNA viruses (106). Some evidence points towards CpGs directly interacting with toll like receptors (TLRs) 3, 7, and 8 and inducing an immune response, however at the cellular level this CpG activation has yet to be observed (107, 108). While it is interesting that CpGs are largely avoided throughout all RNA viruses, and most likely plays a role in avoiding the immune system, it would not be a large enough factor to control the transcriptional machinery and stabilization of the genome.

Finally, the last level of constraint on NSV genomes is the structural properties of the transcribed mRNAs. For optimal translation to occur, the secondary structure of the mRNAs has a direct effect on not only the speed of translation, but also in the resulting protein structure (109, 110). One example occurs in HIV. It has been shown that the secondary structure of mRNA in the gag polyprotein and envelope protein of HIV slows down translation and allows for proper protein folding (111). While this could play a significant role in controlling the translation of

proteins, and influence protein folding, it may also be a level of constraint on the genomic sequence. However, this may be especially important when it comes to the translation of the mRNAs available in positive strand RNA viruses. In NSVs, while this is certainly an evolutionary constraint, we have hypothesized another level of genomic control, all based on the transcription by the vRdRp.

Later in this work, we have put forward a model that considers the interactions between the genomic viral RNA and the nucleoprotein. We postulate that the interactions between the genomic RNA and the nucleoprotein have a direct effect on the processivity of the vRdRp. Since the polymerase must open the nucleocapsid to access the sequestered RNA, there must be some genomic sequences which will have more interactions between the nucleoprotein and the RNA. This would make certain sections of the nucleocapsid harder to access, while making other sections easier. Altering the accessibility of the RNA in certain spots, would influence the overall processivity of the vRdRp. In the work presented in Chapter 4, it was found that just by changing the nucleotide content of the genomic sequence, and not the encoded protein sequence, the activity of the vRdRp drastically decreased. Furthermore, the model is corroborated with a sequential analysis of the genomic content and *in vitro* melting experiments of the nucleocapsid-like particles (NLPs). These experiments point toward genomic base stacking as being the main driving force behind modulating the accessibility of the vRdRp. This paradigm shifting model, represents a completely novel way of thinking about how NSVs modulate their transcriptional/replicational machinery.

1.3.2 Vesicular Stomatitis Virus

In this dissertation, Vesicular stomatitis virus has been chosen as the model pathogen for all the presented investigations. Vesicular stomatitis virus (VSV) has traditionally been used as a

model pathogen for investigating the basic science behind the guiding principles of non-segmented NSVs (112). It is the ideal model system due to its simplicity and transferability when investigating the NSV family. It has a similar architecture to other more pathogenic viruses like Ebola, Marburg, Rabies, and RSV, but is relatively non-pathogenic in humans (112, 113). It's native host is cattle and carrier is the sand-fly, however, it can still infect humans causing mild flu-like symptoms (114). Also, VSV has been shown to be adaptable, being able to tolerate the addition of many different proteins to its genome—making it an excellent candidate for cellular level studies (115, 116). It is for this reason that VSV is chosen for all the studies presented in this work.

Belonging to the family Rhabdoviridae, and closely related to the Rabies virus, VSV is a simple NSV having a genome that is only 11.2 kb encoding for 5 different genes (1). It is a non-segmented NSV and the 5 genes get transcribed sequentially to produce a protein gradient with the first gene being most abundant, and the last gene being least abundant. The first gene encodes for the nucleoprotein (N). The nucleoprotein is responsible for sequestering the genome, serving as the template for transcription and replication all the while sequestering and protecting the RNA. The second protein, known as the phosphoprotein (P), acts as a chaperone protein to bring the polymerase to the nucleocapsid and assists in the formation of the vRdRp. The matrix protein (M) primarily assists in the assembly of the virion, aiding in packaging of the viral proteins to form the classical bullet shape of VSV. The glycoprotein (G) is purely responsible for mediating cell entry through clathrin-mediated endocytosis. Finally, the large protein (L) is the polymerase protein, which is made up of 5 different domains, containing a typical RNA-dependent RNA polymerase (RdRp) domain, a capping domain (to synthesize the 5' caps on mRNA), and also a methyltransferase domain (117). Combined with the N and P-proteins, they form the entire

vRdRp complex (118). Something which is very interesting to note, is that the polymerase protein will not recognize the nascent RNA molecules and be able to carry out its chemistry. It must recognize the nucleocapsid template, chaperoned by the P-protein, in order for transcription to take place (119).

While it is important to have an overview of the function of each protein in VSV, it also important to know the overall structure of the virus (**Figure 1.3C**). It is within these structural properties, from which the functions begin to arise. Starting from the outer layer-working in, VSV is a bullet shaped, enveloped virus. This envelope is made up of a phospholipid bilayer gained while budding off during replication (120). Embedded in this envelope is the glycoprotein. The glycoprotein completely encompasses the outside of the virus and can assume three different conformational states: a native state which is abundant on the outside of the virus at pH 7; an activated hydrophobic state to target the membrane bound receptor (prefusion); and the post-fusion state taking place at low pH after viral entry (121-125). The prefusion structure, allows the VSV glycoprotein to bind to the surface cellular receptor. However, once the virus enters the cells via endocytosis the pH drops to 4. This change in pH mediates a structural rearrangement of the glycoprotein. The prefusion glycoprotein is dimeric, while the low pH post-fusion glycoprotein structurally rearranges to a trimer. The G-protein of VSV is quite unique in that it does not act like a traditional class I or II viral fusion protein—it has properties that mimic both. This has lead researchers to speculate that it could belong to a new category of fusogenic glycoproteins. Recently, it was found that the putative receptor target for the VSV G-protein is the LDL-receptor family. Structural evidence points towards the G-protein targeting two cysteine rich domains, called CR2 and CR3, with identical binding motifs to the different regions (126).

Working our way inward (**Figure 1.3C**), the matrix protein (M) is dispersed around the outer layer of the virion (90). The M-protein is promiscuous in its functions throughout the VSV life cycle. One role is to serve as a structural protein to initiate packaging and budding of the virion. It binds the nucleocapsid protein to initiate the helical packaging of the capsid, while simultaneously stopping viral transcription and replication (127-129). Also, it binds very tightly to the cell membrane and in vivo experiments have shown that even 2M KCl or a pH of 11 cannot disrupt these interactions. This is indicative of its role in the viral life cycle, packaging the contents of the virion into a cell membrane bound vesicle to eventually bud off. This role, in mediating viral packaging and budding has been thought to be facilitated through the M-proteins interactions with Dynamin (130). Dynamin is a GTPase which is responsible for the formation of vesicles for clathrin-mediated endocytosis (131). This would provide the driving force for virion budding, to form a new vesicle. The second role which the M-protein plays during the viral life cycle has to do with the inhibition of various host proteins. It is able to block the host's ability to do translation and transcription (132, 133) and mediates viral evasion of the immune system (127). It has been postulated that the M-protein may suppress the immune system through multiple different modes. One way that the M-protein achieves immune suppression is through inhibition of the Beta Interferon gene (127). The M-protein has been found to be directly correlated with the suppression of Beta Interferon. Suppression of Beta Interferon, would dampen the immune response to VSV infection and allow it to proliferate to nearby cells (134). The M-protein also blocks the host's ability to do transcription and translation. While the methods to which the M-protein achieves this are still highly debated, one way might be through binding to a complex of Rae1 and Nup98 (132). Rae1 is a protein with an uncertain function, but is known to be localized in the cytoplasm, nucleoplasm, and the nuclear rim (135). It can form

many different oligomers and has been found interact with Nup98 (136). Nup98 is a nucleoporin protein responsible for creating a pore for mRNA to be transported outside of the nucleus (137). It was found that the matrix protein readily interacts with the Rae1-Nup98 complex (132). Furthermore, if Rae1 is silenced, host gene expression was not affected, and it increased the cell's ability to resist M-protein induced transcriptional silencing. This indicates that transcriptional silencing is not mediated through the formation of an M, Rae1, Nup98 complex, but through the interaction of the tripartite complex with some unknown moiety. This unknown complex is hypothesized to inhibit host transcription and replication.

When it comes to VSV replication, the M-protein displays a surprisingly diverse range of functionalities. Not only is it involved in the process of viral assembly and eventual budding but is also directly implicated in the virus' ability to evade the immune system. However, perhaps the most interesting and under explored area of research is how the M-protein inhibits the host's transcriptional and translational machinery. Without a thorough understanding of this role, it is almost certain that VSV would not be able to efficiently replicate. This small protein displaying such a wide array of functions is the epitome of viral proteins. Having this wide array of functions, especially for a small virus like VSV, is one of the very important mysteries that must be untangled to fully understand VSV infection.

As we continue to work our way into the virus, the proteins that make up the vRdRp are left. Inside the layer of the matrix protein, is the nucleocapsid protein (N)—this protein will be discussed at length later in this chapter. Seemingly randomly dispersed throughout the bullet and attached to the nucleocapsid are the Phosphoprotein (P) and the polymerase (L) (**Figure 1.3C**). Along with the N-protein, these three proteins form a complex to create the active vRdRp. It is important to note that the vRdRp needs all three proteins to carry out transcriptional chemistry.

The polymerase cannot transcribe the RNA if it is not encapsidated by the N-protein. The overall structure of the vRdRp has yet to be solved, but many functional studies have been done regarding the interactions that dictate this complex (138-143). However, it is without reservation that until the structure of the entire vRdRp can be solved, the basic science behind this important complex will be hard to unravel. Once this structure has been solved, the parameters that regulate this complex will then be able to be actively studied and applied to other pathogenic viruses. However, a lot of information can be gleaned by studying each protein independently.

The P-protein, plays an interesting role in the formation of the vRdRp. It acts as a chaperone for the L-protein to bind the nucleocapsid (144). Sections of the P-protein's structures have been solved both alone and in complex with the N-protein—it is also present in the L-protein structure, but the density is not clear. The P-protein can be thought of having 3 different domains: the N-terminal domain, the Oligomerization domain, and the C-terminal domain (145, 146). Each domain has been structurally characterized separately. Both the N-terminal domain and the C-terminal domain have been solved in complex with the N-protein, while the Oligomerization domain has been solved by itself (141, 145, 147). According to the structural evidence, the N-terminal domain is important for binding what has been called the N0 complex of the nucleocapsid. This complex is the first complex formed before any RNA has been transcribed by the vRdRp (148, 149). This is an important step which occurs only during replication and would represent the first step of replication. The C-terminal domain of P binds the flexible C-terminal loop of the N-protein—far from the encapsidated RNA (141). Upon the C-terminal domain binding, the flexible loop on the N-protein becomes stabilized. Based on the structural evidence, this is believed to be an area where the P-protein is attached to the nucleocapsid. This is important for the formation of the vRdRp. Finally, the Oligomerization

domain is purely responsible for dimerization of the P-protein. Dimerization of the P-protein through the oligomerization domains is important for the processive interactions between the N and L-proteins (145). It is believed that oligomerization of the P-protein potentiates its chaperone activity. Through dimerization, one monomer could interact with the nucleocapsid, while the other interacts with the L-protein.

While P is an important factor in the formation of the vRdRp, it potentially has some control over viral RNA synthesis (150, 151). The P-protein, aptly named, can be phosphorylated in many different positions (152-156). There have been many studies which link phosphorylation of certain regions on the P-protein to viral RNA synthesis (157, 158). One particular study, utilizing mass-spectrometry on the authentic virus, identified Tyr 14 on the N-terminal domain of P to be of vital importance in facilitating viral RNA synthesis only during replication (159). Furthermore, in the same study the authors noted other phosphorylation sites only play a role during transcription. This leads to interesting suggestions of how different phosphorylation patterns within P could impact viral transcription/replication. This would be plausible, as the virus needs various ways to control its viral transcriptional machinery.

The L-protein, carries out all the polymerase activity for VSV. The L-protein is the largest protein in the genome, far surpassing the other proteins in size. For this reason, it also is the protein that has been the hardest to characterize. It wasn't until a couple of years ago (2015) when a Cryo-EM structure was finally solved for the L-protein (117). It was the first full-length polymerase structure to be solved in this family of viruses. The L-protein is versatile and can carry out all the associated polymerase activities, including capping, methylation, and polyadenylation (160-163). The L-protein contains 5 different domains with different functions. It contains the canonical right handed "finger-palm-thumb" structural motif which can be found

in many polymerase complexes (164). This has been called the RdRp domain and is where the transcriptional chemistry occurs. This “finger-palm-thumb” structural motif is surprisingly conserved, even in the heterotrimeric influenza virus polymerase complex. This structural conservation extends from about residue 107-866 (the end of the RdRp domain) in VSV-L and from 415-714 of the PA subunit/8-586 of the PB1 subunit in human influenza B virus (117, 165, 166). This shows that similar mechanisms garner the control of transcription in many of these viruses.

The Capping Domain (A.A. 866-1334) is responsible for capping the 5’ end of the viral RNA. As of now, this domain has no structural homologues and synthesizes the 5’ cap in an unusual way. The host capping mechanism relies mainly on a guanylyl transferase, whereas the chemistry in the capping domain is different. In short, the 5’ end of the RNA is covalently linked to a histidine residue (H1227). This linkage is then attacked by a guanosine nucleotide, capping the RNA and releasing the covalent linkage. This classifies the capping mechanism as a polyribonucleotidyl transferase (PRNTase) (117). Furthermore, there are two different postulated metal binding sites in this domain. Being far away from the active site, they are postulated to bind Zn and are believed to only play a structural role in the protein. Also, this domain is the home to a priming loop. While it is known that VSV L does not need a primer to initiate transcription, it has been postulated that the priming loop could either be used to enhance the fidelity of the polymerase, by sterically slowing it down, and/or by coupling capping to initiation of transcription polymerization.

Moving to the next domain, the Connector Domain (A.A. 1358-1557) appears to be a strictly structural domain. It places the catalytic subunits in the correct position for efficient catalysis (117). Bordered by two different disordered linkers on the N-terminal and C-terminal

sides, this domain is made up of a bundle of eight helices. The Connector Domain, is also postulated to be the place where P may interact with L. To solve the cryo-EM structure of the L-protein, the authors incorporated the N-terminal domain of P during protein purification. However, the authors were not confident of the placement of the fragment in this region but noticed some extra undefined density (117). The authors postulated that this portion of P is responsible for binding to the Connector Domain and stabilizes the L-protein in this structural conformation. This would allow the L-protein to assume a stable structure for the correct placement of the catalytic domains during viral transcription.

The Methyltransferase Domain (A.A. 1598-1892) is able to methylate both the N7 on guanosine and O2' on the conjugate ribose. It does so by transferring this methyl group from the canonical S-adenosyl methionine, the ubiquitous methylation substrate utilized by many different enzymes (167). It has a similar structural configuration with other flavivirus methyltransferases as well as some other methyltransferase enzymes (168-170). Interestingly, the authors stated that there is no tunnel through the protein that connects the methyltransferase domain with the RdRp domain in this conformation (117). However, this would make sense as the structure of the L-protein is not bound to the nucleocapsid. It would be pertinent to predict that upon binding the nucleocapsid, a significant structural rearrangement would occur and cause these domains to come in closer contact. A direct path from the methyltransferase domain to the RdRp domain is needed to produce a viable 5' cap for the mRNA transcripts. However, this tunnel would be useless during replication. This would indicate that there are multiple structural conformations which the L-protein could assume during viral RNA synthesis.

Finally, the last domain of the L-protein is the C-terminal domain. Like the Connector domain, the C-terminal domain likely plays a purely structural role (117). It consists of a bundle

of alpha helices and a 25 residue C-terminal arm. This arm protrudes from the C-terminal domain and loops back towards the RdRp domain and interacts with the previous capping, connector, and methyltransferase domains. This structural feature seems to be conserved among negative stranded RNA polymerases, but the current function is currently unknown (117).

Overall, the L-protein exhibits an incredible array of functions. Much is left to be discovered concerning both the structure and the mechanism of this protein. Moreover, the elucidation of the entire vRdRp complex, would be a break through discovery for understanding how viral RNA synthesis of VSV occurs. Currently it is postulated that the L-protein binds the nucleocapsid and opens three subunits at once, revealing the sequestered RNA (141). During replication, as the RNA is being transcribed it will immediately be encapsidated by an empty nucleoprotein. The polymerase will then open the subsequent subunit while closing the previous one. Obviously, this would mean that there are multiple catalytic structures/rearrangements that must occur for this chemistry to work. This fascinating system is similar to other non-segmented RNA viruses. As evidence, both the nucleocapsid and L-proteins show high sequence identity and similar domain organizations as Rabies, Ebola, Measles, and RSV (117). This would indicate that understanding the basic science behind VSV and its viral RNA synthesis, can be applied to many other important pathogens.

As a model system, VSV helps clarify the guiding principles behind many viral mechanisms. The study of VSV has also led to many potentially breakthrough treatments (171, 172). It is used as the main vector in the highly publicized Ebola vaccine (70), which some vaccination studies have found it to be 100% efficacious. This vaccine is created by adding the glycoprotein of Ebola to VSV, which substantiates a robust immune response in humans. This shows that studying VSV does not only allow one to discover the underlying basic science

behind more pathogenic viruses but can also be adapted towards the development of vaccines. The ability of VSV to be such a highly translatable tool is the reason it is used as the model system throughout this work.

1.4 Literature Review of the Nucleocapsid in Vesicular Stomatitis Virus

As outlined in the previous section, the nucleocapsid is responsible for sequestering the genomic material of the virus (89). Being part of the vRdRp complex, it plays a vital role in viral transcription and replication (118, 141, 160). Since the RNA is sequestered in the nucleocapsid, the L-protein must open the nucleocapsid and abstract the RNA for transcription/replication to occur. In this case, the L-protein cannot carry out its polymerization chemistry on naked RNA alone, it must utilize the protein-RNA template for transcription to occur. Biochemically, this unique system represents an exceptional chance to study the latent interactions that can occur when dealing with transcription from a protein-RNA complex. Unraveling these interactions reveals how nature can apply ubiquitous chemistries in unorthodox ways. Biologically this system yields the opportunity to examine more clinically relevant questions like: Can the nucleocapsid be a good target for a novel, transferrable antiviral? And what are the constraints placed on NSV viral RNA synthesis? When coupled with a multi-faceted understanding of biochemistry, answering these questions can lead to some important answers. As the nucleocapsid of VSV is focus of this work, it is important to understand what is already understood in the literature.

The structure of the nucleoprotein (N) of VSV in complex with RNA was solved by our lab back in 2006 (89). This was the first structure of its kind to be solved and was published simultaneously along with the nucleocapsid protein structure from Rabies (85). It was in these two seminal papers that revealed the intricate intermolecular interactions which guide the

assembly and encapsidation of the viral RNA. The N-protein is made up of two distinct lobes: the N-terminal lobe and the C-terminal lobe (**Figure 2.1** in Chapter 2). Sandwiched between these lobes is a pocket where the RNA is sequestered. Each monomer of the N-protein interacts with 3 other monomers of the nucleocapsid protein (89). It does this using a protruding N-terminal arm and flexible C-terminal loop, both of which are important for both assembly and RNA encapsidation. On the N-terminal lobe is the N-terminal arm. This N-terminal arm is anchored into an adjacent C-terminal lobe through multiple interactions. Furthermore, to further enhance these interactions, the C-terminal loop interacts with a previous C-terminal lobe and an N-terminal arm. This creates a tightly interweaved intermolecular assembly between the nucleocapsid proteins.

Located between the N-terminal lobe and C-terminal lobe, the genomic RNA sits in a C-shaped binding pocket. The RNA encapsidated by the N-protein is highly structured. Each subunit encapsidates 9 nucleotides. With a slight helical structure, the encapsidated RNA closely resembles a single stranded B-form RNA. Furthermore, it has a unique base stacking pattern which is repeated throughout each subunit (89). Nucleotides 1-4 base stack with each other, while nucleotide 5 base stacks with 7 and 8. Nucleotide 6 is flipped in the opposite direction and does not base stack with any other nucleotides, while nucleotide 9 is labile and is able to base stack first nucleotide in the adjacent subunit. This base stacking pattern is sequence independent. The structures of the nucleocapsid were solved with all of the of polynucleotides available: poly(rA), poly(rG), poly(rC), and poly(rU) (173). Comparing these structures to the randomly encapsidated RNA from the recombinantly expressed protein showed that the global structure of both the RNA and protein do not change with sequence (173). While the structural properties of both the RNA and protein do not change, it was noticed that the occupancy of the RNA changed

slightly with differing nucleotides. This would suggest that the nucleocapsid modulates nucleotide binding in some unforeseen way. Additionally, a mutational study of the amino acids present within the RNA binding pocket revealed complete abolishment of RNA binding. Both biochemical and crystallographic evidence show that the mutant S290W cannot encapsidate RNA (174). The bulky W side chain encompasses the pocket and blocks RNA from being encapsidated while preserving the intricate intermolecular interactions relevant to assembly. This structure indicates, and another structure which was solved without RNA (173), that the nucleocapsid assembly is independent of RNA binding. This means that assembly of the nucleocapsid occurs with or without RNA present. Also, once the nucleocapsid is assembled it retains the ability to encapsidate RNA.

There are two structures that have been solved of the nucleocapsid protein in complex with a fragment of the P-protein: the C-terminal domain of P and the N-terminus of P (141, 147). Both complexes are unique and presumably represent different stages of transcription and replication. The structure of the C-terminal domain of P (PCTD) and the N-protein (N/PCTD) can only occur when the N-protein is fully assembled. The PCTD binds at a site formed between two different subunits on the top of the C-lobe (141). It binds a part of the C-terminal loop and α-helix 13 on the N-protein. PCTD is comprised of 5 α-helices and an antiparallel β-turn. Interestingly, the overall structure of both the PCTD and N-protein do not change after forming the complex. Overall, 24 hydrogen bonding interactions aid in forming the complex with 19% of the total surface area of PCTD contributing to the creation of the N/ PCTD complex. This structure represents a significant step in understanding the pertinent interactions associated with the formation of the vRdRp. Since P acts as a chaperone to bring the L-protein to the nucleocapsid, understanding these interactions through a direct method allows for the design of a

model of replication. Using this structure, one can predict that the PCTD could be largely responsible for transiently anchoring the L-protein in an optimal position for access to the sequestered RNA in the nucleocapsid. In doing so, the PCTD needs to be able to simultaneously both bind and release the nucleocapsid as the polymerase moves down the genome (141). While this structure may not necessarily be relevant to both transcription and replication, it is unknown as to which part of the replication cycle this structure represents. However, it provides an important molecular snapshot of the viral life-cycle and yield important information about the formation of the vRdRp.

Another sought after structure is the N0 structure. To be best defined, the N0 structure is an empty N/P complex formed during the initial step of replication. This structure has been proposed to act as a switch, triggering the vRdRp to change from transcription to replication. The hypothesis is that after a significant protein build up, the N0 complex forms and binds to the vRdRp (148, 149, 175). The act of binding initiates a change from transcribing protein, to replicating the entire genome. It would first form +cRNA and then -vRNA. During transcription, the product of the vRdRp is mRNA, which is then translated and packaged into proteins. However, during replication the nascent RNA needs to be immediately encapsidated. This is because the polymerase complex must then re-copy the entire genome to complete viral replication—the RNA produced cannot go on to produce protein. Thus, the formation of the N0 complex a critical step in the viral life cycle. In the literature, there is a proposed N0 structure (147). Multiple studies have confirmed that a truncated N-terminal arm, when co-expressed with full-length P will yield an interesting RNA free 2P-N complex, but this has yet to be fully structurally characterized (112, 147). In the putative N0 structure the truncated N-protein is in complex with a portion of the N-terminus of the P-protein (P60). Like the previous data this N0

structure is also RNA free. In solution, this complex forms monomeric heterodimers (1:1 P:N), while *in crystallo* it forms the traditional decameric complex. The binding mode of P60 sits between the N-subunits and blocks a part of the RNA cavity. This fragment also occupies the same space as the N-terminal arm. Using Small Angle X-ray scattering and Size Exclusion Multi-Angle Light Scattering to compare the crystal structure to the structure in solution, the authors found that these complexes are structurally similar. Furthermore, NMR data confirms the crystal structure configuration but suggests that there is a high degree of solvent exchange on the first 17 residues of P60. This indicates that N-terminal portion is conformationally flexible but prefers the more stable *in crystallo* orientation (147, 176). This structure shows that P does not only chaperone L to the nucleocapsid for transcription to take place, but also brings the monomeric nucleoprotein to the nascent RNA being produced by the vRdRp. While this structure certainly reveals a molecular snapshot of a possible N0 complex, more work throughout the field needs to be done to further investigate its role.

While the complexes of the N and P-proteins have revealed a lot about the molecular underpinnings which control RNA encapsidation, very little is known about the molecular architecture of the tripartite complex that forms the vRdRp. Functional mutational studies have attempted to clarify this (177), and are certainly worth exploring, but the actual contacts cannot be mapped until the structure is solved. Utilizing the knowledge gained through the elucidation of the nucleocapsid structure, an aim of this work is to functionally characterize a selected region which was hypothesized to be critical for the formation of the vRdRp. Mutational studies when guided by structure, can be a great guide to reveal mechanism and architecture in the context of a viral infection. Unfortunately, without a complete structure of the vRdRp conclusions drawn from these types studies must be validated using other biochemical methods. However, once the

structure of the vRdRp is solved, the functional information present in the literature becomes the foundation for determining mechanism.

In this family of viruses, the nucleocapsid protein is paramount to the viral life cycle. Playing the crucial role as the gatekeeper to the genome, means that it has many functions. Throughout this work many differing aspects of this protein are thoroughly examined and a background of the existing literature is actively needed. As it comes to the structure-function relationships of this protein there are still many mysteries yet to be resolved by the literature. This work is meant to take a step in the direction of further clarifying the basic mechanisms which guide NSV replication.

1.5 Research Objectives and Synopsis

This dissertation aims to investigate the basic chemical properties controlling the biological processes surrounding the nucleocapsid protein of vesicular stomatitis virus and how these processes regulate the activity of the vRdRp. Furthermore, across all NSVs, the nucleocapsid exhibits stunningly high levels of structural conservation. In this light, all the presented studies have been designed to be applicable in the context of the wider family of NSVs. Utilizing a highly interdisciplinary approach, with techniques taken from Biochemistry, Molecular Biology, Biophysical Chemistry, Bioinformatics, and Virology, this dissertation can be viewed through a three-tiered lens:

1. Development of Novel Antiviral Therapies
2. Further the understanding of the architecture of the viral RNA dependent RNA polymerase through mutational studies
3. Elucidate the genetic constraints on the RNA Polymerase which control viral RNA synthesis

Examining the first lens, the development of antiviral therapies is not only good for eradicating diseases but also can be useful to aid in the understanding of basic biological functions. Furthermore, using rational drug design and preexisting classes of compounds, specific antivirals can sometimes be found with little off target effects and toxicities. In the case of this study, a polyamide was adapted to target the encapsidated RNA in VSV. Polyamides are well-known class of compounds that can specifically target the minor groove of DNA. They are well studied with relatively low toxicities and can be targeted to specific sequences of DNA (61, 62). Also, they have been found to be an effective antiviral agent against HPV and are currently under further development (63, 65). In this study a polyamide was found that can target the RNA in the nucleocapsid of VSV and inhibit viral RNA transcription. While this compound is specific for VSV, this is the first known case in the scientific literature of a polyamide targeting viral RNA. Also, the knowledge gained in our approach can be used to develop other polyamides against a multitude of NSV infections.

Exploring the second lens, understanding the interactions that drive the assembly of the vRdRp in VSV is still an ongoing research topic. While every protein in VSV has either had the entire or a portion of the structure solved by Cryo-EM or X-ray Crystallography, the architecture of the vRdRp and how the proteins interact with each other remains elusive. To obtain an active vRdRp, three proteins need to be present: the Nucleocapsid, the Phosphoprotein, and the Polymerase (or L-protein). To fully understand the transcription and replication processes, the interactions guiding both the assembly and function of the vRdRp need to be fully elucidated. Without a crystal/Cryo-EM structure, alanine scanning mutation studies represent a viable biochemical method to tease out intricate interactions between the proteins. In this study, utilizing previous structural information, mutations around a proposed RNA “access gate”

proved to abolish almost all vRdRp activity. These mutations did so without interrupting the global structure of the nucleocapsid, indicating direct interaction by disrupting either the formation or activity of the vRdRp. Finally, these mutations were introduced into the VSV genome and viral passages commenced. After multiple passages, compensatory mutations were found on the Phosphoprotein and the L-protein, which restored the vRdRp activity. This is indicative that these amino acids are pertinent towards creating an active vRdRp complex. While solving a structure of the vRdRp is the goal, the results presented in this study has significantly furthered our understanding of the interactions that are necessary for the formation of an active vRdRp.

The third lens investigates how the vRdRp self-regulates its processivity. For the vRdRp to carry out its transcriptional chemistry, it first must access the RNA by opening the nucleocapsid. Thinking about this structurally and chemically, one can imagine a region in which the interactions between the nucleoprotein and the RNA are strong and/or weak. These strong/weak regions would have a significant impact on the ability of the polymerase to access the RNA and carry out transcription. This would mean that the processivity of the vRdRp would change based upon the sequence content of the genome. To validate this, the sequence content of the genome, and not the encoded protein content, was changed and the activity of the vRdRp was assessed. It was found that changing the sequence content could actively modulate the processivity of the vRdRp. Utilizing our structural knowledge of the nucleocapsid it was proposed that base stacking interactions, arising from the structural properties of the RNA in the nucleocapsid, might play a role in this stability. To assess this, a bioinformatics approach was used to analyze the sequence content of both the wt and altered gene sequences. In doing so, it was found that the ability of purine/pyrimidines to base stack with each other significantly

changed the properties of the sequence. Finally, to experimentally validate that purine/pyrimidine base stacking alters the stability of the nucleocapsid, we assessed the nucleocapsid containing either a purine (poly(rA)) or pyrimidine (poly(rU)) by a thermal shift assay. It was found that purines significantly increase the stability of the nucleocapsid, while pyrimidines decrease the stability. In doing so, we propose that it is the purine/pyrimidine base stacking content which can modulate the stability of the nucleocapsid. In changing this stability, it also changes the processivity of the vRdRp. This finding is particularly important in understanding how NSVs regulate protein production and various forms evolutionary constraints. The contents of this study could be directly applied to the development of vaccines and targeted antivirals towards the NSV family.

This three-tiered approach has given a complete picture of the structure function relationships that guide the biological function of the nucleocapsid protein in VSV. A thorough understanding of these relationships has allowed for the design of a novel antiviral compound, the elucidation of important amino acids involved in viral RNA transcription, and a paradigm shifting approach of the understanding in how NSVs modulate viral RNA synthesis through the interaction of the genome with the nucleoprotein. This understanding of the basic science behind the driving forces of how the nucleocapsid is intimately involved in the viral replication life cycle, represents a shift in the understanding of its functional properties. These three studies represented in this work epitomize an interdisciplinary approach towards illuminating the basic chemical and structural principles that guide the biology of NSVs. The underlying theme of this work is the utilization of these techniques and how they can bring about a broader understanding of not only VSV, but towards the underlying mechanisms that control the NSV life cycle. In doing so, the knowledge gained in this dissertation can be directly applied towards more

pandemic pathogens, in the hope of bettering the human condition and life-span by preventing emergent diseases from taking hold and potentially saving lives.

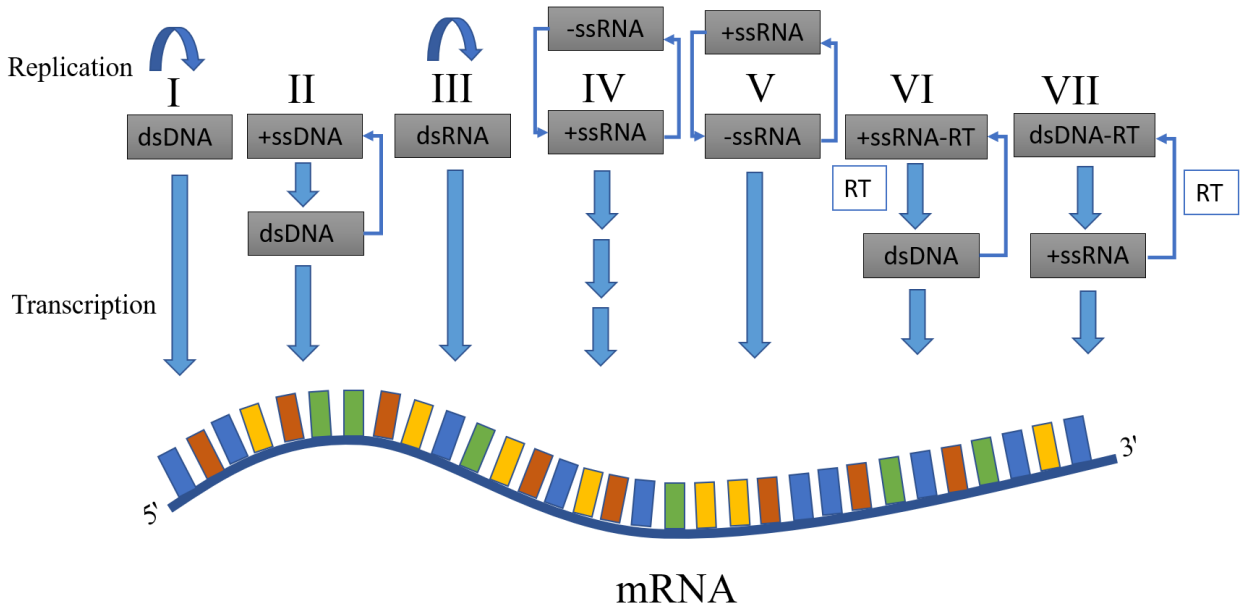


Figure 1.1 Baltimore Classification System

Schema of the Baltimore classification system devised by David Baltimore, which groups viruses based on their method of replication.

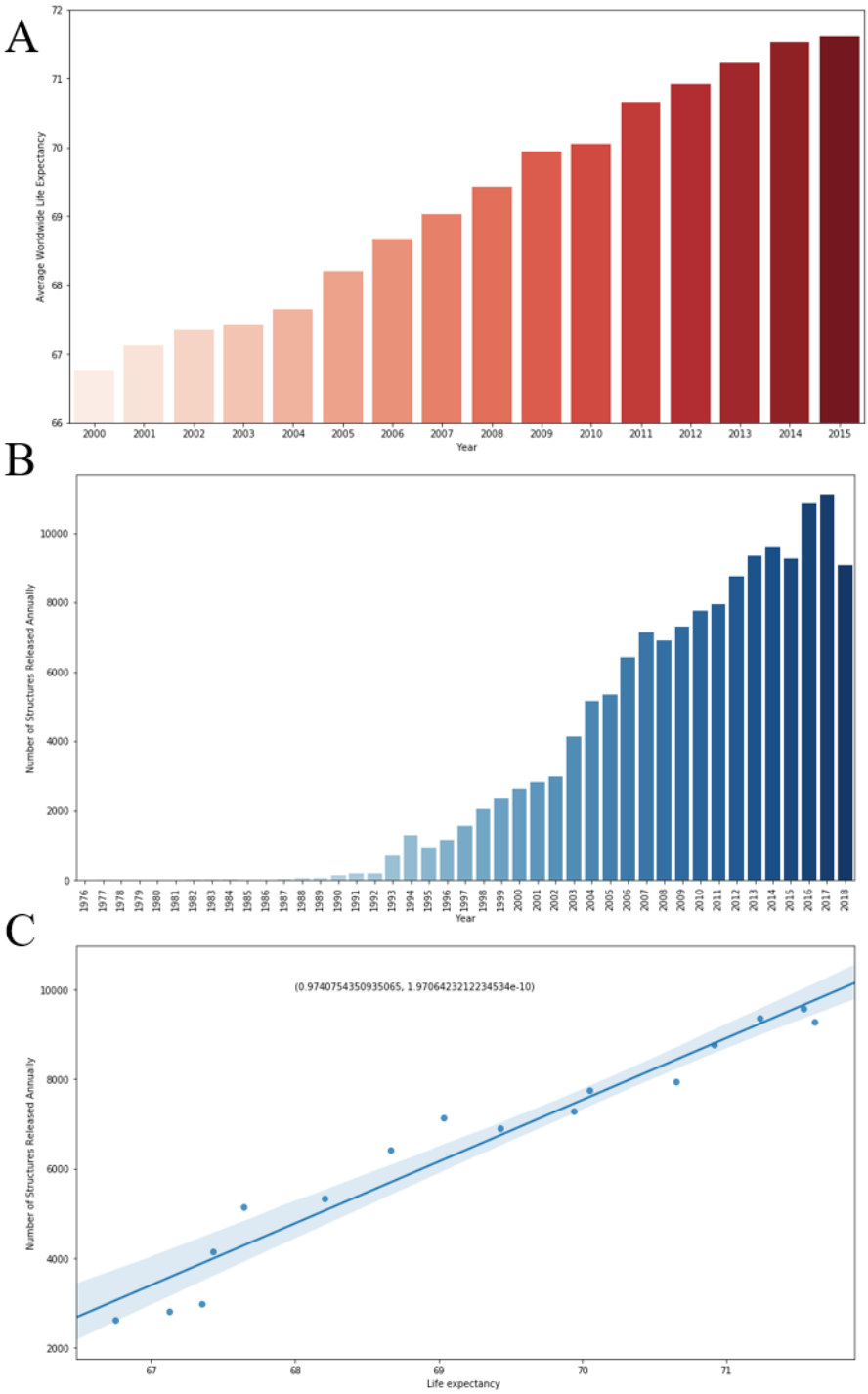
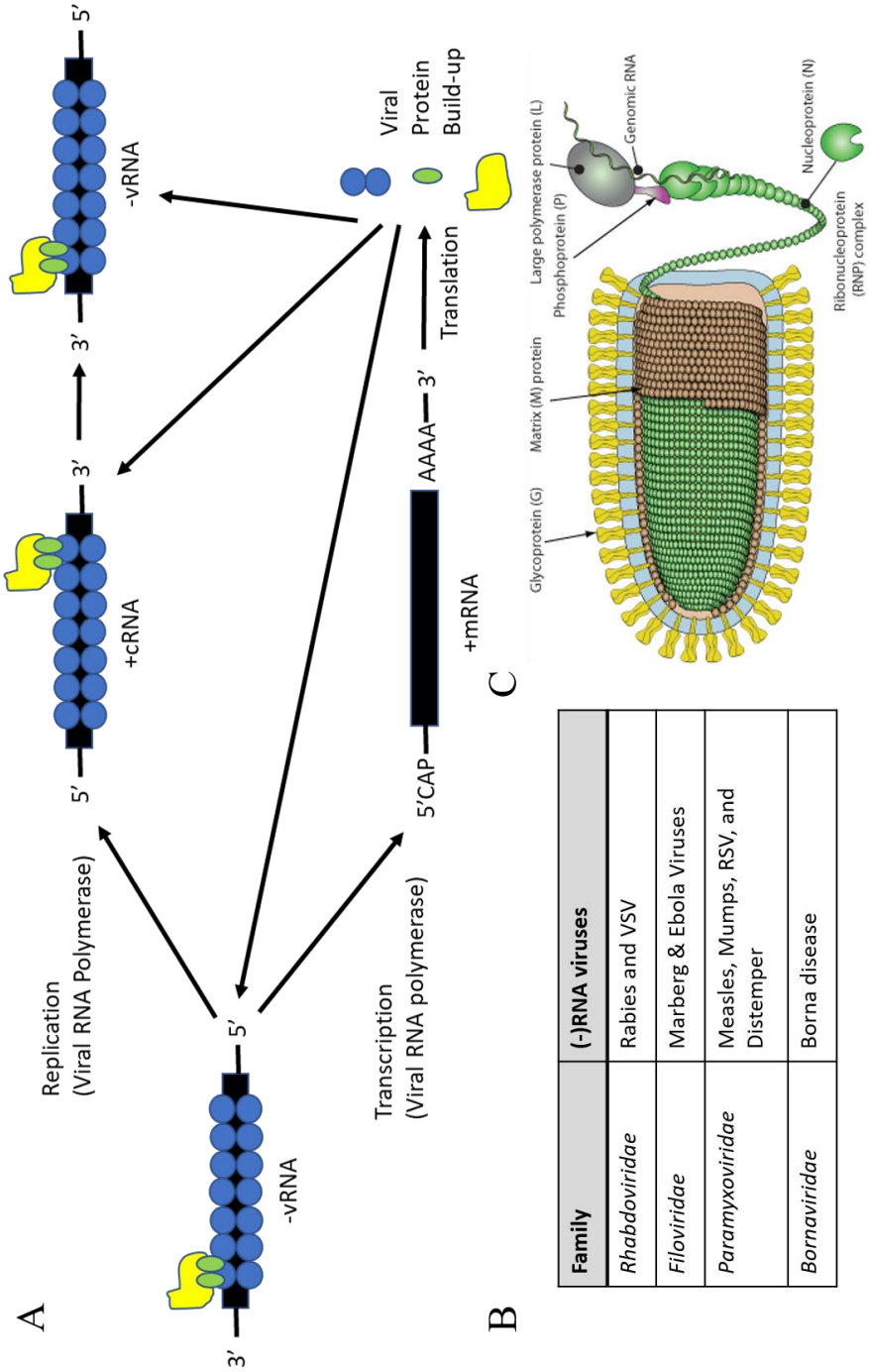


Figure 1.2 Life Expectancy and Published Protein structures.

Life Expectancy and Published Protein Structures. A. A bar graph showing the average life expectancy increasing by nearly 5 years from 2000-2015. B. The number of deposited structures in the Protein Databank increasing every year since the early 1990's. C. There is a strong correlation between Life Expectancy and the number of Protein Structures Deposited. While it is not causation, it speaks the importance of science.



Family	(-)RNA viruses
Rhabdoviridae	Rabies and VSV
Filoviridae	Marberg & Ebola Viruses
Paramyxoviridae	Measles, Mumps, RSV, and Distemper
Bornaviridae	Borna disease

Figure 1.3 Negative Strand RNA Virus Overview

Negative Strand RNA Virus Overview. A. Method of replication utilized by Negative Strand RNA Viruses. B. A succinct table of some commonly known Negative Strand RNA Viruses and their Family of classification. C. Cartoon drawing of Vesicular Stomatitis Virus used with permission. ViralZone: https://viralzone.expasy.org/21 SIB Swiss Institute of Bioinformatics.

2 A POLYAMIDE INHIBITS REPLICATION OF VESICULAR STOMATITIS VIRUS BY TARGETING THE RNA IN THE NUCLEOCAPSID

 [Journal of Virology, Volume 92, Issue 8, April 2018, e00146-18, DOI: 10.1128/JVI.00146-18]

2.1 Abstract

Polyamides have been shown to bind double-stranded DNA by complementing the curvature of the minor groove and forming various hydrogen bonds with DNA. Several polyamide molecules have been found to have potent antiviral activities against papillomavirus, a double-stranded DNA virus. By analogy, we reason that polyamides may also interact with the structured RNA bound in the nucleocapsid of a negative strand RNA virus. Vesicular stomatitis virus (VSV) was selected as a prototype virus to test this possibility since its genomic RNA encaspsidated in the nucleocapsid forms a structure resembling one strand of an A-form RNA duplex. One polyamide molecule, UMSL1011, was found to inhibit infection of VSV. To confirm that the polyamide targeted the nucleocapsid, a nucleocapsid-like particle (NLP) was incubated with UMSL1011. The encapsidated RNA in the polyamide treated NLP was protected from thermo-release and digestion by RNase A. UMSL1011 also inhibits viral RNA synthesis in the intracellular activity assay for the viral RNA-dependent RNA polymerase. The crystal structure revealed that UMSL1011 binds the structured RNA in the nucleocapsid. The conclusion of our studies is that the RNA in the nucleocapsid is a viable antiviral target of polyamides. Since the RNA structure in the nucleocapsid is similar in all negative strand RNA viruses, polyamides may be optimized to target the specific RNA genome of a negative strand RNA virus, such as respiratory syncytial virus and Ebola virus.

2.2 Introduction

Many pathogenic viruses are negative strand RNA viruses (NSVs), such as respiratory syncytial virus (RSV), and parainfluenza viruses (e.g. PIV3), which cause severe infections in children. For instance, RSV accounts for up to 80% of hospitalization due to acute respiratory infections in children (77). Treatment options are limited and no effective vaccines are currently available, which creates an urgent unmet need for novel antiviral medications for respiratory illness in children and adults. Furthermore, deadlier NSVs include rabies virus and Ebola virus, for which no effective antiviral drugs are available (178-180).

One common feature of all NSVs is that they encode their own RNA-dependent RNA polymerase (vRdRp) in the genome. During transcription/replication, the vRdRp must access the structured RNA genome sequestered in the nucleocapsid and will not recognize the free RNA genome as a template for RNA synthesis. This makes the nucleocapsid a prime target for broad antiviral therapies. The crystal structure of a nucleocapsid-like particle (NLP) was first solved for vesicular stomatitis virus (VSV)(89) revealing that the nucleocapsid protein (N) forms a polymer with extensive intermolecular interactions to package the genomic RNA (174). The N-protein has a deep cleft between its N-terminal and C-terminal lobes where the genomic RNA is sequestered (**Figure 2.1A**). When the genomic RNA is packaged in the nucleocapsid, it assumes a unique structure (**Figure 2.1B**). If we count the nucleotides from 5’ end to 3’ end, 9 nucleotides are covered by each N subunit. The bases of nucleotide 1 to 4 are stacked with the bases facing the exterior of the cleft, whereas bases of nucleotide 5 and nucleotides 7-9 are stacked with the bases facing the interior of the cleft. The base of nucleotide 6 does not stack with other bases. The base stacking of nucleotides 1-4 or 5,7-9 are curved. The orientation of the ribose and the

phosphate groups do not have a regular conformation either. The bases of VSV RNA are more parallel to each other in nucleotides 1-4, than 5, 7-9.

One such class of drugs that have been shown to target nucleotide moieties are polyamides. Several polyamides have been developed as inhibitors of gene expression and other biological effectors that bind double stranded DNA (dsDNA) in a sequence-specific manner (58-62, 181-183), and other polyamides have also been developed as broad-spectrum antivirals for papillomaviruses and polyomaviruses precisely because they do not show sequence-specific binding properties (63-65, 184-188). Others have also since discovered nonspecific polyamide effects (189, 190). Polyamides have not been cytotoxic to HeLa (191) or other mammalian cells in these reports, in one case showing SI values of at least 2000 before reaching compound solubility limits (65). Within the G nucleotides of a DNA helix, imidazole (Im) moieties of polyamides can act as an H-bond acceptor of one hydrogen from the exocyclic NH2 as predicted by Dickerson (192). Furthermore, pyrrole (Py) residues and β-alanine residues (β) typically bind to A, T, and C through H-bond donation from amide bond NH groups to acceptor N or O groups in the minor groove.

Since polyamides have been shown to target structured nucleotide moieties and form hydrogen bonds, such as the structure found in the nucleocapsid-bound RNA of VSV, we tested a group of polyamides for possible antiviral activities against VSV (see **Table 2.1**). One molecule, UMSL1011 (**Figure 2.2**), was found to inhibit VSV replication. Further experiments found that UMSL1011 could protect the encapsidated RNA from thermo-release, digestion by RNase A, and inhibit vRdRp mRNA synthesis. Finally, the crystal structure of UMSL1011 bound to the VSV NLP was solved and it was found to directly target the encapsidated RNA. Our results indicate a novel use of polyamides and can potentially be developed as antiviral

drugs against other NSVs which cause widespread diseases. This opens a new avenue for drug discovery to effectively control NSV infection.

2.3 Results

2.3.1 Antiviral Screen

A plaque assay for VSV infection was set-up for antiviral screens. Six different polyamide compounds (**Table 2.1**) prepared previously were included. When inhibitor solutions at 20 μg/mL (130 μM) concentration of polyamides in 1% DMSO were added, UMSL1011 inhibited VSV infection (**Figure 2.3**). This result confirmed that a polyamide compound could inhibit infection by VSV. The antiviral activity by UMSL1011 is specific because other polyamide compounds did not inhibit VSV infection at this concentration. Inhibition of VSV infection by UMSL1011 was repeated with different concentrations of UMSL1011 (**Figure 2.3**). At <10 μg/mL (<65 μM) , UMSL1011 inhibits plaque formation by VSV infection.

2.3.2 Melting Curve Analysis

In order to verify that UMSL1011 is targeting the nucleocapsid, Differential Scanning Fluorimetry (or a Thermal Shift Assay (TSA)) was carried out. More specifically we wanted to test if UMSL1011 could delay the thermal release of the RNA from the NLP. Two different fluorescent dyes were selected to target the encapsidated RNA and protein, respectively. For RNA, SYBR Safe was used to stain DNA/RNA in electrophoresis gels, because it works by intercalating DNA/RNA and then fluorescing. Since the bases of the RNA inside NLP are stacked, and the local RNA concentration is high inside the NLP as well, clear fluorescence was observed when SYBR Safe was incubated with NLP (**Figure 2.4**). To approximate the temperature when the RNA is dissociated from the NLP, we call this Tfree, the 1st derivative graph was calculated for the rate of fluorescence changes. Most importantly, the peak in the 2nd

derivative of the melt-curve defines the temperature point at which the rate change is at the maximum. We name this temperature as Tfree to indicate the RNA release from the NLP. To confirm our findings, SYPRO Orange was used to detect protein melting. The melting of the protein in the NLP is correlated with the release of encapsidated RNA because RNA can stabilize the protein in NLP (173, 174). We have previously observed RNA release from NLP when temperature was increased (173). Since the melting of the protein roughly follows the release of the RNA, as defined by Tfree, ΔTfree can be used to investigate the stability of the encapsidated RNA in the NLP.

Using this assay, it was shown that when 200 μg/mL (1.3 mM) of UMSL1011 was incubated with NLP, Tfree was shifted upward by ~2 °C (68.86 vs 66.82 °C) (**Figure 2.4**). We interpreted this observation of Tfree shift as UMSL1011 binding and stabilizing the RNA in NLP. A shift of 2 °C in Tm is typical when an organic compound binds a protein in vitro (193). On the other hand, Tfree was only shifted by ~0.3 °C from an inactive polyamide molecule UMSL1013 (**Figure 2.4**), suggesting that UMSL1013 does not stabilize the encapsidated RNA in the NLP. Interestingly, UMSL1011 has a slight effect on protein stability, showing a -0.9 °C shift when compared to the DMSO control (69.86 °C w/ UMSL1011, and 70.75 °C w/DMSO alone) (red lines in **Figure 2.4**). Furthermore, the unfolding of the protein begins after RNA is released from the NLP in all cases.

2.3.3 RNA protection

To further verify our melting curve analysis findings, RNA from NLP was exacted by phenol extraction. The freed RNA was subjected to RNase A digestion in the absence and presence of 200 μg/mL (1.3 mM) UMSL1011 in 1% DMSO. There was no protection of the freed RNA from RNase A digestion by UMSL1011 (**Figure 2.5A**). This observation is

consistent with our hypothesis that UMSL1011 could only bind the structured RNA in NLP. Digestion of RNA encapsidated by the NLP was also carried out. Previously, we have shown that RNA inside NLP could be digested by RNase A (173). In this case, digestion of RNA in the NLP was protected by UMSL1011 (**Figure 2.5B**), whereas the inactive UMSL1013 allowed for RNA digestion (**Figure 2.5B**).

2.3.4 Evaluation of cytotoxicity

Previous studies showed that UMSL1011 and its analogs did not show cytotoxicity up to 200 μM (65). UMSL1011 were evaluated for cytotoxicity by MTT assays using CellTiter 96® from Promega. It was found that even up to 650 μM of UMSL1011, it exhibited very little cytotoxicity (**Figure 2.6C**). This is indicative of a low probability of off-target effects and is consistent with previous data.

2.3.5 Minigenome Assay

To further test the specificity of UMSL1011 to inhibit VSV viral RNA synthesis, a commonly used minigenome assay/RT-qPCR was adapted to quantify mRNA transcripts (194-197). The N gene with three stop codons after the initiation Met codon was used as the reporter gene to encourage no bias. With this assay, UMSL1011 was shown to inhibit viral synthesis of mRNA for the reporter gene. The raw CT values for the qPCR assay are shown in **Figure 2.6A** and a normalized ΔCT value are shown in **Figure 2.6B**. The ΔCT was calculated by comparing the house keeping gene, β-actin, to the CT value of the mRNA. This was then normalized to the Mock ΔCT values allowing the calculation of a relative % inhibition of transcription for UMSL1011. Compared to the DMSO control transfection, UMSL1011 at 20 μg/mL (130 μM) inhibited transcription of mRNA by 54.5%, whereas the DMSO compared to the Mock about 8.2%. It is worth noting that the p-value of >0.05 for comparing the Mock to the DMSO control

was not statistically significant. On the other hand, the p-value for the UMSL1011 sample compared to the Mock/DMSO control was $p<0.001$, showing a high level of confidence in statistical significance. Also, to corroborate our analysis with the well known 2-ΔΔCT analysis **Appendix A.1** show these values. These further confirm the pattern that is seen with our minigenome analysis.

2.3.6 Crystal Structure

In order to fully define the interactions of UMSL1011 with the structured RNA in the NLP, the crystal structure of UMSL1011 bound NLP was determined by X-ray crystallography (see **Table 2.2** for statistics). The electron density corresponding to Py-3, Im-4, γ-5, and Py-6 of UMSL1011 (see **Table 2.1** and **Figure 2.2**) was found in the NLP (**Figure 2.7A**). Due to resolution limitations and randomly encapsidated RNA when the NLP was purified from E. coli, the RNA was first modeled as a poly-U chain, as done previously (89, 173). In total UMSL1011 forms seven hydrogen bonds with the RNA and two hydrogen bonds with the N-protein (**Figure 2.7B**). The electrostatic surface of UMSL1011 based on APBS calculations confirmed the hydrogen bond potential of each UMSL1011 group (**Figure 2.7C**). This confirmed that the methyl group in Im-4 is polarized so their hydrogens may form hydrogen bonds with a proton acceptor, as found in (198-200). The amide nitrogen and the carbonyl group directly preceding the Im-4 each make a single hydrogen bond with the base of nucleotide 9. The three polarized methyl hydrogens coming off the Im-4 each form a hydrogen bond with the two riboses of nucleotide 7 and 8, and the phosphate between nucleotides 7 and 8. Im-4 also participates in weak π-π stacking with the base of nucleotide 9, further anchoring the Im-4 in place. Furthermore, the amide nitrogen and carbonyl group after Py-3 each forms a hydrogen bond with the base on nucleotide 3 and 4, respectively. The carbonyl group directly preceding Py-6 forms a

hydrogen bond with Arg-312, seemingly anchoring the γ-5 turn in place, and the methyl group on Py-6 interacts with the backbone carbonyl on Gln-318. The side of UMSL1011 that interacts with RNA is more positively charged, whereas the opposite side is mostly negatively charged. Unlike other polyamide interactions that would normally target the minor groove of DNA (201), UMSL1011 stretches across the structured RNA by making contacts with both the bases and the backbone (**Figure 2.7D**). Interestingly, UMSL1011 fits in a pocket created by the interface between two N subunits and the encapsidated RNA (**Figure 2.7A**).

For the genomic RNA in the VSV nucleocapsid, there could be any nucleotide sequence in each position that interacts with UMSL1011. To define the optimal sequence for binding with UMSL1011, we modeled each base at the RNA position in the UMSL1011-NLP complex. The possible hydrogen bonds between UMSL1011 and NLP are summarized in **Table 2.3**. This model suggests that 5'AXXCUXXGG3' may be the preferred sequence motif to bind with UMSL1011. The specific base is designated for that nucleotide position based on optimal hydrogen bond formation with the interacting moiety of UMSL1011. Survey of the genomic sequences found that there are 18 of this motif in the negative strand and 6 in the plus strand, which represent all potential binding sites for UMSL1011.

To investigate whether UMSL1011 induces any structural changes, we aligned the A chain of the UMSL1011 bound structure with the A chain of PDB 2GIC—the original structure of the nucleocapsid. These crystals were grown under the same conditions as 2GIC. There is almost no change in the N subunit structure where UMSL1011 is bound (**Figure 2.8A**). However, globally the overall NLP ring has tightened significantly. The diameter of the original NLP is 46.6 Å, measured from the C-terminus to C-terminus of two opposite N subunits in the

ring, whereas the UMSL1011 bound NLP has a diameter of 44.3 Å (**Figure 2.8B**). This ring tightening of 2.3 Å can be corroborated with our TSA and RNA protection data.

2.4 Discussion

Many pathogens are negative strand RNA viruses, including rabies virus, respiratory syncytial virus and Ebola virus. Currently, there are no antiviral drugs targeting viral RNA synthesis by its unique viral RNA-dependent RNA polymerase (vRdRp). A large body of structural studies on the nucleocapsid of NSVs revealed a common mechanism of NSV viral RNA synthesis (202). During viral transcription and replication, vRdRp recognizes the nucleocapsid and gains access to the genomic RNA sequestered in the nucleocapsid to use it as the template. vRdRp must induce conformational changes in the nucleocapsid to release the genomic RNA for initiation and elongation during viral RNA synthesis. This provides a unique opportunity for antiviral inhibitors. Since the nucleocapsid is the template for viral RNA synthesis, viral replication will be blocked by a compound that binds the nucleocapsid.

By plaque assays, six different polyamide compounds were screened for inhibitory activities against VSV infection (**Figure 2.3**). One compound, UMSL1011, was found to inhibit VSV infection. UMSL1011 was further confirmed to specifically inhibit viral RNA synthesis using an intracellular vRdRp minigenome assay (**Figure 2.6**). It is noted that UMSL1011 reduced viral plaques by about 50% at concentrations between 65-130 μM.. This is further corrorborated by the minigenome assay, which shows a reduction of 54.5% in the mRNA levels synthesized by the vRdRp when 130 μM of UMSL1011 was added. This would indicate that UMSL1011 could not inhibit viral RNA synthesis very efficiently in this concentration range, allowing for a small amount of virus to continue through replication. Furthermore, the reduction in activity of viral RNA synthesis is not due to off target effects of UMSL1011. Previously, it

was shown that UMSL1011 had little effect on the gene regulation in the apoptotic, dna repair, and cell cycle pathways (184).While this is important to note, it allows one to use UMSL1011 as an existing model to design more active polyamide based inhibitors.

To clearly define the target and reveal the mechanism of UMSL1011 inhibition, a melting curve and RNA protection studies were carried out on recombinantly expressed NLP. Due to the possible sequence specific manner in which UMSL1011 interacts with the NLP and to account for the randomly encapsidated RNA from E. coli, the concentration of UMSL1011 was increased for all in vitro experiments on the NLP. By systematically increasing the temperature, structural changes are expected to occur in the nucleocapsid. Two fluorescent dyes were used to monitor the structural changes. One dye, SYBR Safe, intercalates between stacked bases of RNA. At the beginning, the fluorescence of SYBR Safe was high when incubated with NLP, suggesting tight binding to NLP. When the temperature continued to increase, the fluorescence of SYBR Safe started to drop indicating reduction in RNA binding, but the fluorescence of SYPRO Orange, the other dye we used to monitor protein unfolding, did not change very much. At a critical temperature, the fluorescence from the RNA released began to increase, indicating that that the RNA had already been released. To define a more common point after RNA release, the 2nd derivative was utilized and named Tfree. The observation of thermo-release of RNA from NLP was consistent with our previous studies of RNA digestion (173). However, we interpreted it as that the RNA sequestered in NLP was thermally released and named this temperature as Tfree. Past this temperature, the fluorescence of SYPRO Orange began to increase rapidly and a second Tm was observed, suggesting unfolding of the nucleocapsid protein. This indicates that the encapsidated RNA can stabilize NLP. When the melting curve measurement was repeated in the presence of UMSL1011, Tfree was increased by ~2°C (**Figure**

2.4), indicating binding of UMSL1011 stabilizes NLP and prevents thermo-release of sequestered RNA. When an inactive polyamide UMSL1013 was added, Tfree was not increased. Subsequently, an RNA protection assay was carried out to confirm that UMSL1011 binds only the RNA structure in NLP (**Figure 2.5**). If UMSL1011 was able to interact with the unstructured RNA extracted from the NLP, we would have seen distinct fragmentation of the RNA strand after RNase digestion. However, when the RNA in NLP was extracted, it was completely digested by RNase A in the absence or presence of UMSL1011, suggesting no interactions between free RNA and UMSL1011. On the other hand, the RNA in NLP may be digested by RNase A when the temperature was raised to release RNA (173). When NLP was incubated with UMSL1011, the RNA in NLP was protected from RNase A digestion, indicating that UMSL1011 binds the RNA in NLP, further indicating that UMSL1011 stablizes the NLP. The slight shift in the migration pattern is due to slight nibbling of exposed ends of the RNA by RNase. At the same time, the inactive UMSL1013 did not protect the RNA in NLP from RNase A digestion. The in vitro studies further argue that the polyamide UMSL1011 binds the RNA in the nucleocapsid to prevent it from serving as a template for RNA synthesis by vRdRp.

The co-crystal structure of UMSL1011 bound to the NLP further validates the in vitro and biochemical results. Importantly UMSL1011 was found to interact with both the structured encapsidated RNA and the nucleocapsid protein (**Figure 2.7**). UMSL1011 binds the NLP in a manner not previously observed in polyamides. This unique binding pattern could be optimized to target more pathogenic NSVs. Not only is there an intricate hydrogen bonding network between UMSL1011 and the RNA/protein (**Figure 2.7B**), but also binding UMSL1011 effects the NLP globally causing a tightening of the NLP ring by 2.3 Å (**Figure 2.8B**). Furthermore, a preferred UMSL1011 binding motif could be identified in the genomic RNA sequence (**Table**

2.**3**). Taking all into account, the inhibition of viral RNA synthesis by UMSL1011 could be a two-fold process. The tightening of the nucleocapsid would make it harder for the polymerase to access the encapsidated genome, slowing the polymerase down or causing the complex to fall apart completely. Furthermore, UMSL1011 directly interacts with the RNA causing an impassable blockage in the path of the polymerase complex.

To our knowledge, this is the first time that the encapsulated genome of an NSV has been directly targeted by an antiviral compound. Furthermore, this is the first evidence of a polyamide targeting and inhibiting the RNA synthesis of an NSV. Other NSVs exhibit similar RNA structures with stacked bases found inside the nucleocapsid, including rabies virus, respiratory syncytial virus and measles virus (85, 86, 203). Such structures may also be expected in other NSVs, including Ebola virus and Hantavirus because the structure of their nucleocapsid proteins and the capsid assembly are very similar (204-206). While UMSL1011 may be specific for VSV, it is possible that other polyamides can target other NSVs in a specific manner and would be subject of future studies. Our proof-of-concept studies on nucleocapsid binding and inhibition of VSV infection by UMSL1011 set a precedent for design of polyamide molecules as antiviral agents against other NSVs.

2.5 Methods

2.5.1 Antiviral Screen

VSV (Indiana strain) stocks were prepared in BHK-21 cells (173). Monolayer HeLa cells were cultured in DMEM supplemented with 10% FBS, and grown to full confluence in 12-well plates. After washing with DPBS, VSV was dispensed over the cells and allowed to be absorbed for one hour in 37 °C, 5% CO2. The excess virus inoculum was washed away and the cells were overlaid with DMEM containing 0.8% agarose. After 48 hrs, the cells were fixed with 4% formaldehyde

and the agarose overlay was removed. The plate was stained with 0.1% crystal violet, and the plaques were counted.

2.5.2 Thermal Shift Assay

Prior to carrying out the TSA, the isolation of recombinantly expressed NLP was purified according to the protocol as previously established by our lab (144). All NLPs were checked by UV/Vis to verify the 260/280, ensuring RNA encapsidation, and SDS-page to verify protein purity. To test the efficacy of the compound in the Thermal Shift Assay, a final concentration of 200 μg/mL (1.3 mM) of UMSL1011 or UMSL1013, which corresponds to a final concentration of 1% DMSO, was incubated with 25 μM (1.2 mg/mL) of NLP. To ensure binding to the NLP three rounds of incubation at 42°C for 15 minutes with subsequent incubation on ice for 15 minutes each were carried out—this was also carried out with the DMSO control. Stocks solutions 100X were prepared in the reaction buffer (50 mM Tris pH 7.5, 300 mM NaCl) for the SYBR Safe and SYPRO Orange dyes. Each TSA reaction was carried out in 20 μL volumes which contained final concentrations of 20 μM of NLP, 1% DMSO or 180 μg/mL of UMSL1011/1013 dissolved in DMSO, and a 10X concentration of the dyes. The 20μL reactions were prepared in triplicates and placed into a MicroAmp Fast 96-well plate. A QuantStudio3 was used to carry out the TSA with a gradient of 0.025 °C/s from 25-95 °C. The Savitzky-Golay method was used to smooth the raw data and then 1st and 2nd derivative analysis was done to identify the Tm and Tfree, respectively.

2.5.3 RNA protection

RNA was extracted using Trizol reagent from Ambion by Life Technologies following the manufacturers protocol. After RNA extraction, the RNA was subjected to RNase treatment (final concentration of 15 μg/mL) both with 1% DMSO and 200 μg/mL (1.3 mM) of UMSL1011 for

30 mins and 1 hour respectively. Samples were then loaded on a 2% agarose gel and visualized with SYBR Safe. The RNA encapsidated by the NLP was digested with 1 mg/mL of RNase for 15 or 30 minutes at 42 °C. Prior to RNase digestion, 200 μg/mL of compounds/1% DMSO was added to the NLP and the cycling of temperature was carried out in the same manner as the TSA. Each reaction was quenched by using the Trizol reagent and subsequent extraction of the RNA. The samples were then loaded on a 2% agarose gel containing SYBR Safe and further visualized. Gel Analyzer 2010a was used to quantify the intensity of the bands on the gels, for further analysis.

2.5.4 Evaluation of Cytotoxicity

HeLa cells were grown to confluence in 100 mL of DMEM supplemented with 10% Fetal Bovine Serum (FBS) in 96-well. MTT assays were carried out following the protocol from the manufacturer of CellTiter 96® (Promega, Madison, Wisconsin, USA).

2.5.5 Minigenome Assay

The minigenome assay was designed and carried out as described previously (194-197). We constructed the minigenome to transcribe a non-translatable N gene by adding three stop codons to the beginning of the gene. The minigenome and the helper plasmids were co-transfected and either compound (final concentration 20 μg/mL (130 μM)) or 1% DMSO was added to the media. RNA was extracted using the RNeasy RNA Extraction Kit (Qiagen), using the manufacturer's protocols 48 hours post transfection. Immediately after RNA extraction, reverse transcription was carried out to make the cDNA using the primers 5'-CCAGATCGTTCGAGTCGTTTTTTTTTTTTTTTTTCATTTGTCAAATTCTGACTTAG-3', for the N mRNA transcripts, and 5'-AGCACTGTGTTGGCGTACAG-3', for β-actin, using M-MLV reverse transcriptase from Invitrogen following the manufacturer's protocol. After cDNA

synthesis, qPCR was carried out with PowerUp SYBR Green Mastermix following the manufacturer's instructions for concentrations and cycling protocols. In short, 10 ng of cDNA was used as a template and 500 nM of both primers 5'-GTTGAATGGCTCGGATGGTTC-3' and 5'-CCAGATCGTTCGAGTCGT-3' for the mRNA transcript of N and 5'-AGAGCTACGAGCTGCCTGAC-3' and 5'-AGCACTGTGTTGGCGTACAG-3' for the β-actin transcript. 40 cycles of PCR were carried out and a subsequent melting curve to check for transcript homogeneity. The primer only (polymerase plus primers), the no primer (template plus polymerase), and the no reverse transcriptase reaction showed no discernable CT values when doing the experiment—the fluorescence was under the threshold to discern a CT value. The background fluorescence is taken into account when setting the baseline for the qPCR reactions, utilizing all of the controls as well, thus only using concentrations of the template yielding CT values greater than 15 are used—so one can get an accurate baseline to set the CT value. Data were then analyzed using the ΔCT method for relative mRNA quantitation.

2.5.6 Structure

Crystallization was carried out as previously described with minor modifications prior to crystallization to facilitate UMSL1011 binding (89). Otherwise, everything else was kept as previously. Prior to crystallization, UMSL1011 (1 mg/mL (6.5 mM)) was incubated with the NLP in the same manner as the thermal shift assay. In short, three rounds of incubation at 42 °C for 15 minutes followed by subsequent incubation on ice for 15 minutes was carried out. Hanging drops were set up immediately after incubation and crystals appeared after 7-10 days. Data was collected at the Advanced Photon Source on SER-CAT 22-ID with an oscillation angle of 0.3 °. Data was processed with the XDS-package and molecular replacement was carried out (PDB code: 2GIC as a search model) using maximum-likelihood procedures in PHASER with

the Phenix Suite (207, 208). The structure was then refined using phenix.refine in the Phenix Suite and model building/ligand placement was done in COOT (208, 209). The restraints for the UMSL1011 fragment were generated using eLBOW in the Phenix Suite from a SMILES string (208). All the figures were prepared with PyMol (210). Electrostatic surface potentials of UMSL1011 were generated using APBS and visualized using UCSF Chimera as described previously (211, 212).

2.5.7 Chemistry

Polyamides were synthesized on ABI433A or CSBio 136XT peptide synthesizers using Boc methodology and extensive incorporation of dimer building blocks as previously described (185, 213-218). Boc-β-alanine-PAM resin was obtained from Peptides International and building blocks were from A Chemtek or were synthesized in-house. Dimethylamine-free DMF was from ABI (now Fisher). UMSL1028, UMSL1011, UMSL1055, and UMSL2115 were synthesized, purified and described as previously (65, 184, 216). Compounds were cleaved off the resin with N,N-diaminopropylamine, leading to Dp derivatives, or with bis(aminopropyl)methylamine, leading to Ta derivatives. Preparative reverse phase chromatography was carried out on a Phenomenex C18 Luna column in MeOH/water with 0.1% TFA, and the TFA salts were isolated by lyophilization. All compounds were analyzed for purity by two orthogonal HPLC and HPLC/MS methods: Reverse Phase C12 on a Phenomenex Jupiter column in an MeCN/water gradient (HPLC/MS) and a Phenomenex Synergi Polar-RP column with ammonium formate running buffer for the orthoganonal HPLC. Exact masses were determined by ESI, and 500 MHz or 600 MHz 1H NMR spectra were obtained. Typically, 13C and various 2-D NMR spectra were also obtained. All chemical characterization are present in the Supplemental Materials.

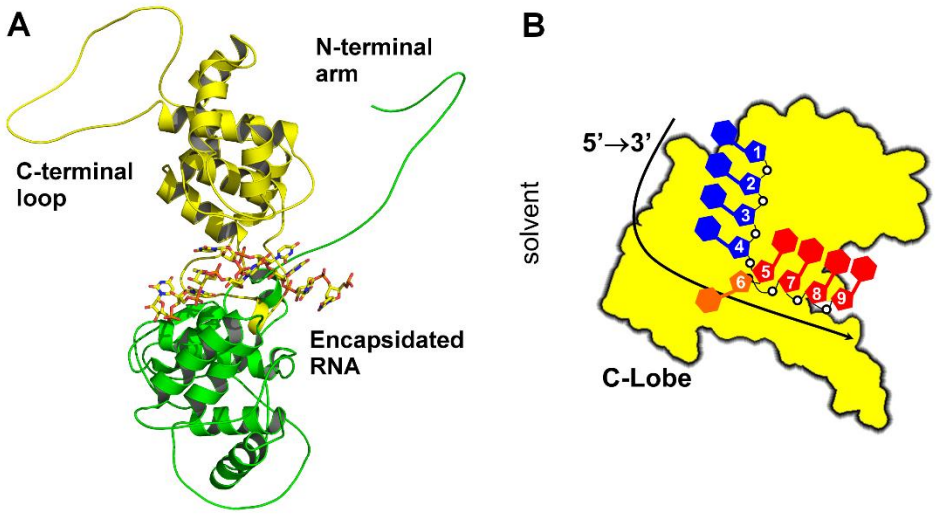


Figure 2.1 RNA Structure in VSV NLP.

RNA Structure in VSV NLP. (A) Ribbon and stick drawings of VSV N in complex with RNA. The N-terminal lobe is shown as ribbons in green, and C-terminal lobe in yellow. The N-terminal arm and the C-terminal loop elements are labeled. RNA structure in the center is shown as a stick model. (B) Cartoon diagram of a single subunit in the VSV NLP from the perspective of the N-lobe. This shows the RNA base-stacking that occurs within the NLP. All drawings were created with PyMol (210).

Table 2.1 Polyamides used in antiviral screen, with their building blocks defined

Compound	N-terminus	2	3	4	5	6	7	8	9	10	C-terminus	Activity
UMSL1011	dIm	Py	Py	Im	γ	Py	Py	Py	Py	β	Dp	active
UMSL1013[#]	dIm	Py	Py	Py	γ	Py	Py	Im	Py	β	Dp	inactive
UMSL1028[§]	dIm	Py	Py	β	Py	Py	Py	y				
(cont.)	Py	Py	β	Py	Py	Py	Py	β	Ta			inactive
UMSL1055	dIm	Im	Py	Py	γ	Py	Im	β	Py	β	Dp	inactive
UMSL2115	fPy*	Im	β	Im	γ	Py	Py	β	Py	β	Ta	inactive
UMSL2082	dIm	Im	Py	γ	Py	Py	Py	β	-	-	Ta	inactive

*dIm indicates an N-terminal n-methyl imidazole lacking an amino group (des-amino-N-methylimidazole-2-carboxylate). Py is derived from 4-amino-2-carboxy-N-methylpyrrole; Im is derived from 4-amino-2-carboxy-N-methylimidazole. β is derived from β-alanine; γ from γ-aminobutyric acid. Dp is derived from N,N-dimethylpropylamine; Ta from bis(aminopropyl)methylamine. #was previously reported and is inactive against HPV(22, 23). §UMSL1028 is a 16-residue polyamide active against HPV16, 18 and 31. *fPy is derived from 4-formamido-N-methylpyrrole-2-carboxylate.*

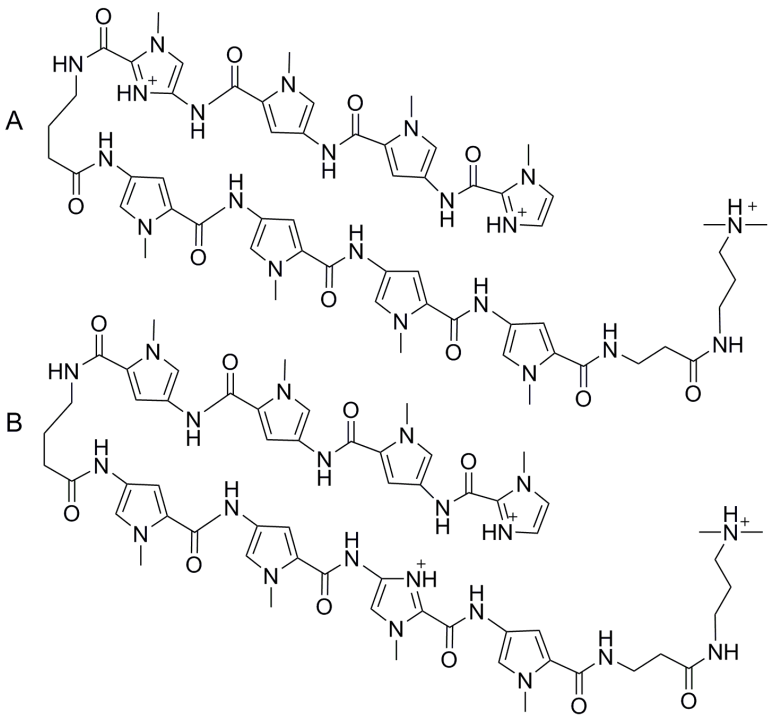


Figure 2.2 Structure of Polyamides Used

Structure of Polyamides Used. A) The cation of UMSL1011 as isolated as a tris(TFA) salt. (B) The cation of UMSL1013 as isolated as a tris(TFA) salt.

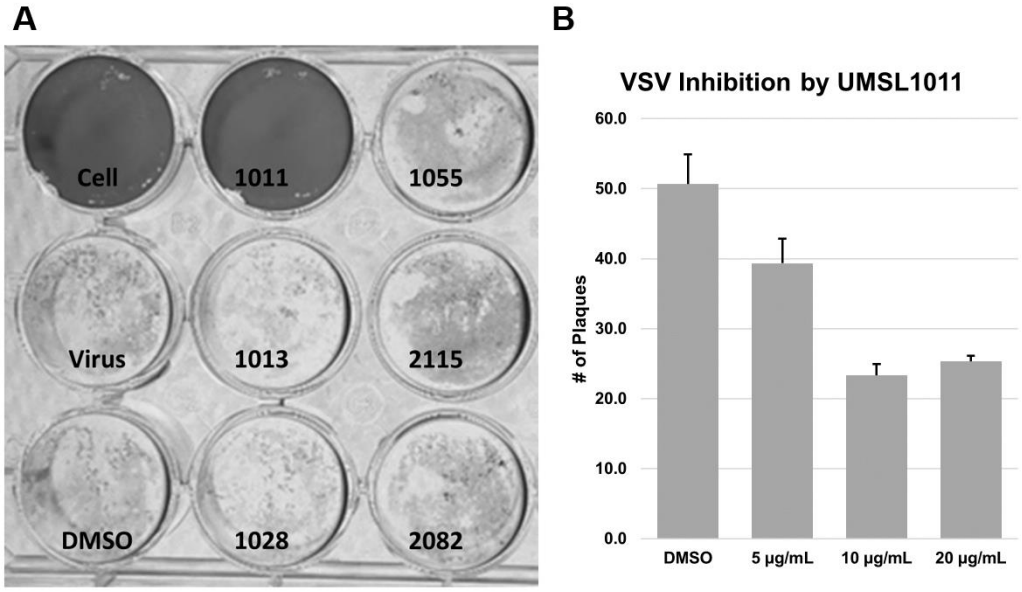


Figure 2.3 VSV plaque assays.

VSV plaque assays. (A) The initial screen of six polyamide compounds (20 µg/mL). Controls of uninfected cell, medium only and 1% DMSO were included. The compound names were labeled for each well. (B) After virus adsorption for 1 hr, UMSL1011 at different concentrations was added to the agarose overlap. DMSO was used as the solvent negative control. Experiments were done in triplets and the plaques were counted at 18 hr postinfection.

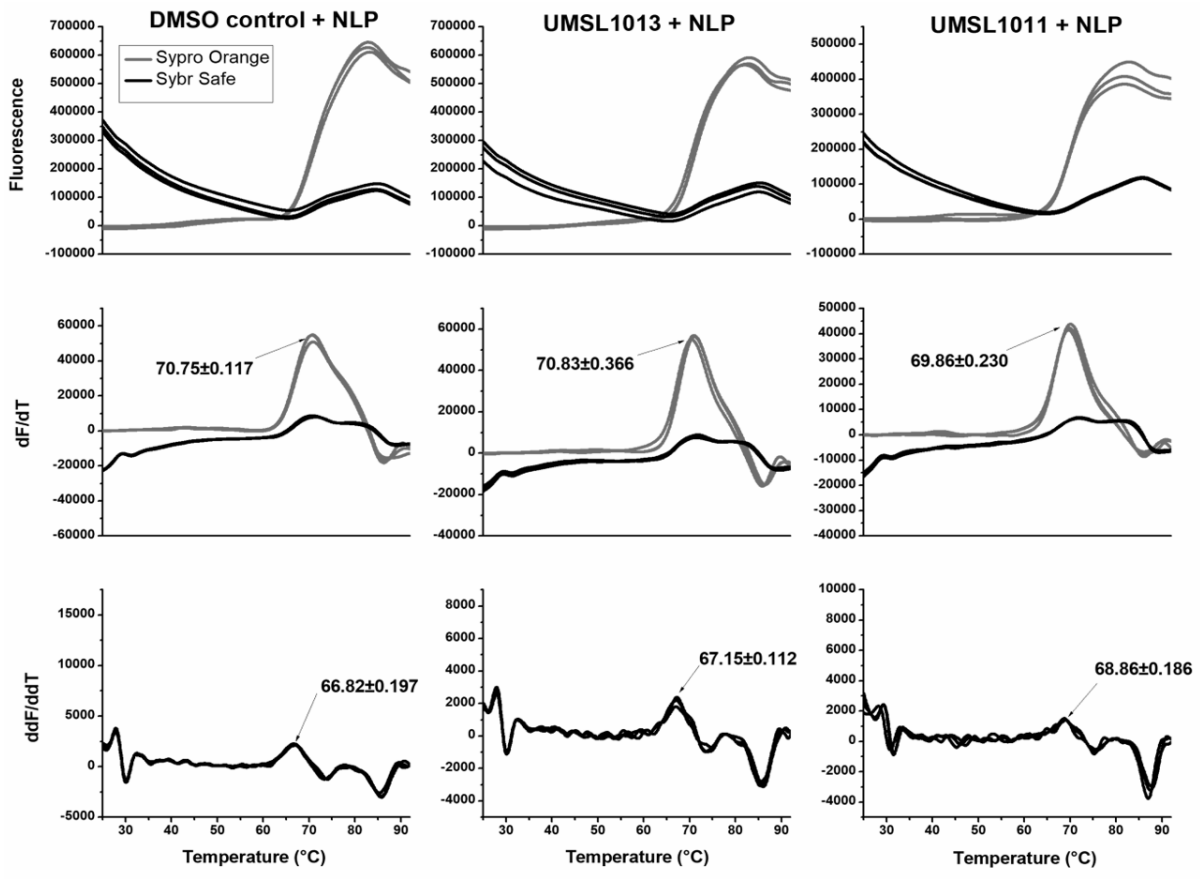


Figure 2.4 Melting curve analysis.

Melting curve analysis. VSV NLP was incubated with 200 µg/mL (1.3 mM) USML1011 (right column), 200 µg/mL UMSL1013 (middle column), and in 1% DMSO (left column). NLP was at a concentration of 1.2 mg/mL in 50 mM Tris, 300 mM NaCl, pH 7.5. Both SYBR Safe and SYPRO Orange were added to the NLP sample (Blue and Red trace, respectively). The temperature was scanned at 0.025 °C/s in QuantStudio 3 instrument. The first derivative and the second derivative graphs are presented. Tfree (= 68.9 °C) (equivalent to Tm) is determined from the second derivative graph. The left column shows the results from the same experiment and presentation as the right column, except that NLP was incubated with 1% DMSO. Tfree (= 66.8 °C) is determined from the second derivative graph. To have a negative control, the Tfree measurements were repeated with an inactive compound UMSL1013 (central column). Tfree was relatively the same in comparison with DMSO control. All errors are shown as the standard deviation.

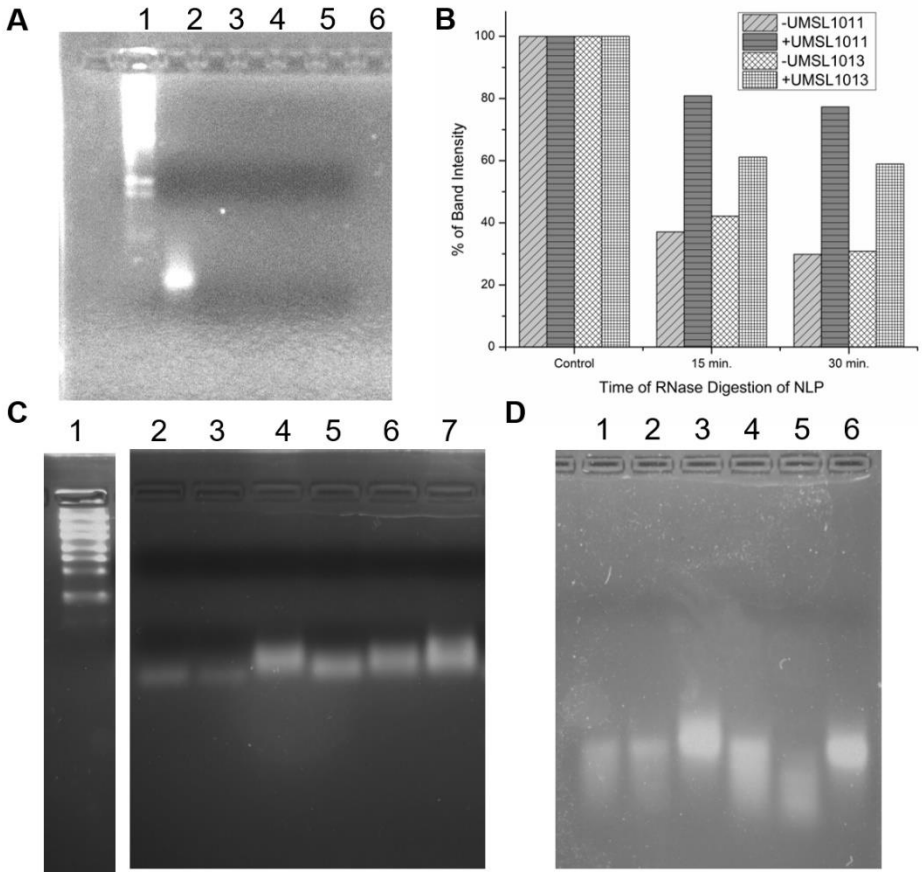


Figure 2.5 RNA protection study.

RNA protection study. (A) RNA freed from NLP (Lane 2) was incubated with 15 µg/mL RNase A with (Lanes 3 and 4) and without (Lanes 5 and 6) 200 µg/mL (1.3 mM) UMSL1011 at 22 °C for 30 (Lane 2 and 4) and 60 min (Lane 3 and 5), respectively. RNA markers (Lane 1) are included. (B) A bar graph quantifying the drop in intensity of each band in the gels shown in panels in C and D. Both +/-UMSL1011 and +/-UMSL1013 are shown compared to each undigested control, which is set at 100%. The program GelAnalyzer 2010a was used to quantify the brightness and overall peak area of each gel and normalized on a percentage based on the control. (C) NLP was incubated with 1 mg/mL RNase A at 42 °C without (Lanes 2-4) and with (Lanes 5-7) 200 µg/mL (1.3 mM) UMSL1011. Lane 2 and 5 were 30 min digestion, and Lane 3 and 6, 15 min, respectively. Lane 4 and 7 had no RNase A. Lane 1 is RNA markers (200 b as the lowest band vs 90 b RNA from NLP). (D) A control study with an inactive compound. 200 µg/mL UMSL1013 (Lanes 4-6) was incubated with NLP (Lane 4, 15 min/Lane 5, 30 min). Lane 6 has no RNase A. Lanes 1-3 corresponds to the same reactions in the presence of 1% DMSO.

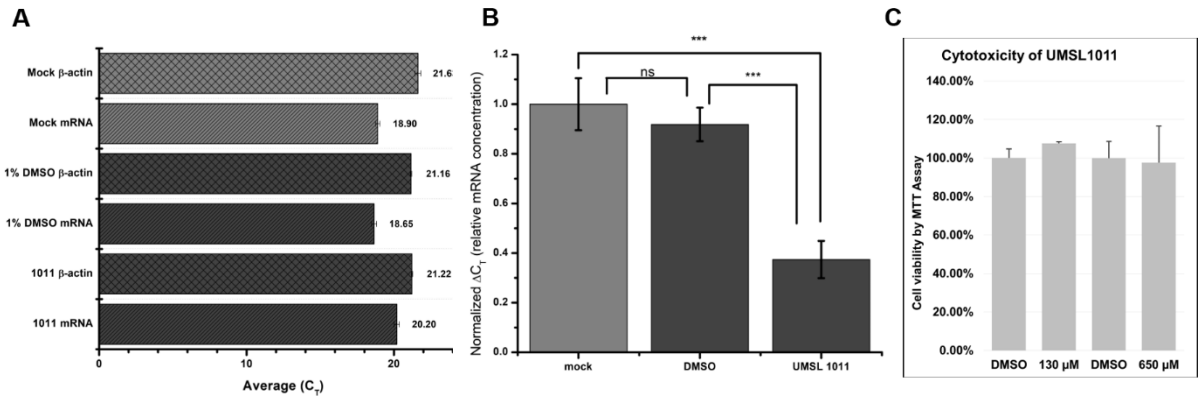


Figure 2.6 Intracellular vRdRp assays and Cytotoxicity Study

*Intracellular vRdRp assays and Cytotoxicity Study. (A) Raw C_T values of mRNA transcripts from the minigenome assay and reporter gene, β-actin. Each C_T value is taken as an average of three qPCR reactions and error bars represent the standard deviation. (B) Normalized ΔC_T values to the Mock, set as 1. Each ΔC_T was calculated by subtracting the experimental C_T value to its subsequent β-actin C_T and the error bars represent the standard deviation. This shows a significant decrease in mRNA transcripts when UMSL1011 is added. *** indicates a $p < 0.001$ and ns indicated a $p > 0.05$ (C) Cytotoxicity studies. Monolayers of HeLa cells were incubated with 130 μM and 650 μM UMSL1011, respectively, for 25 hr. 1% DMSO was used as the solvent control. Cell viability was measured by MTT assays. Experiments were carried out in triplets and error bars representing the standard deviation.*

Table 2.2 Crystallographic Statistics

Wavelength	1
Resolution range	45.39 - 3.46 (3.584 - 3.46)
Space group	P 2 21 21
Unit cell	75.48 165.48 234.46 90 90 90
Total reflections	168290 (16809)
Unique reflections	38347 (3334)
Multiplicity	4.4 (4.4)
Completeness (%)	92.22 (86.26)
Mean I/sigma(I)	6.79 (2.70)
Wilson B-factor	100.3
R-merge	0.1526 (0.6546)
R-meas	0.1718 (0.7406)
R-pim	0.07654 (0.3363)
CC1/2	0.99 (0.742)
CC*	0.997 (0.923)
Reflections used in refinement	36198 (3334)
Reflections used for R-free	1999 (184)
R-work	0.2655 (0.3160)
R-free	0.3107 (0.3572)
CC(work)	0.912 (0.837)
CC(free)	0.868 (0.775)
Number of non-hydrogen atoms	17466
macromolecules	17417
ligands	49
Protein residues	2086
RMS(bonds)	0.003
RMS(angles)	0.76
Ramachandran favoured (%)	86.09
Ramachandran allowed (%)	9.71
Ramachandran outliers (%)	4.2
Rotamer outliers (%)	10.07
Clashscore	21.47
Average B-factor	121.78
macromolecules	121.77
ligands	124.71

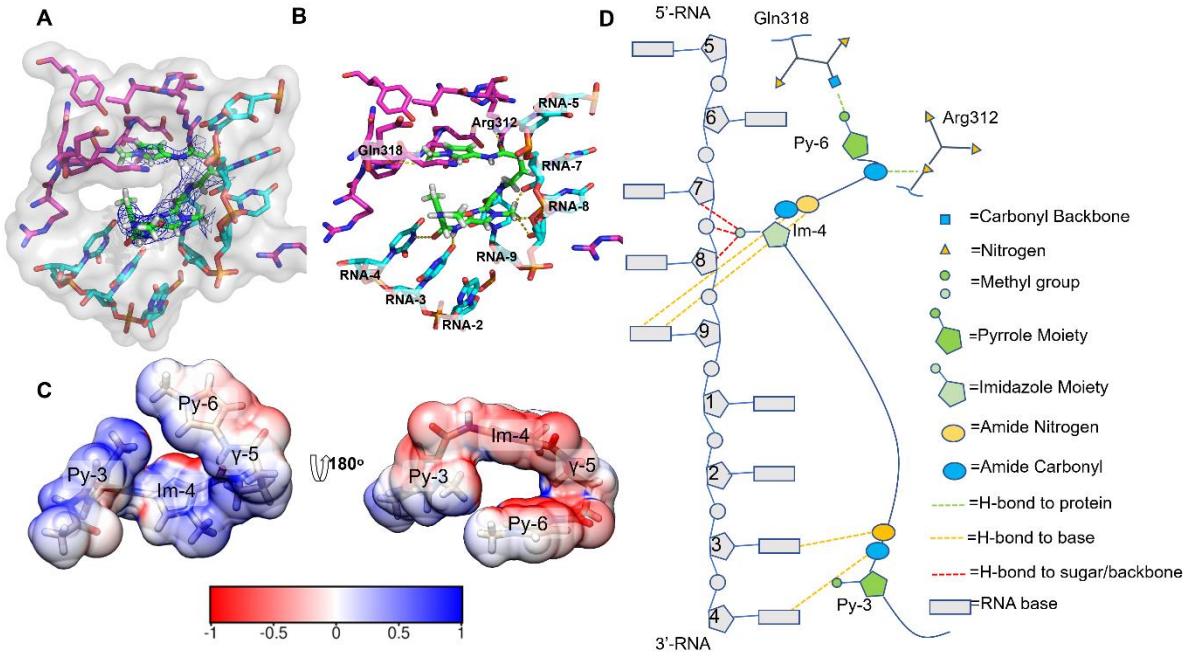


Figure 2.7 Crystal Structure of UMSL 1011 bound to the NLP.

Crystal Structure of UMSL1011 bound to the NLP. A) 2mFo-dFc map (shown in blue) at 1 sigma of UMSL1011 (shown in green) bound to the NLP. The magenta sticks are nearby amino acids of the NLP, while the RNA is modeled in cyan. The transparent surface reconstruction shows the pocket for which UMSL1011 binds. (B) Hydrogen bonds made by UMSL1011 (shown in green) to both the encapsidated RNA (cyan) and the protein (magenta). (C) Electrostatic surface potential of UMSL1011 calculated by APBS and visualized in Chimera . The colored bar below shows the relative charge with red being electron dense and blue being electron deficient. (D) A two-dimensional cartoon outlining the hydrogen bonds between UMSL1011 and the RNA(polyrU)/protein. Protein contacts are shown with Gln318 amide carbonyl backbone, and the Nε of Arg312.

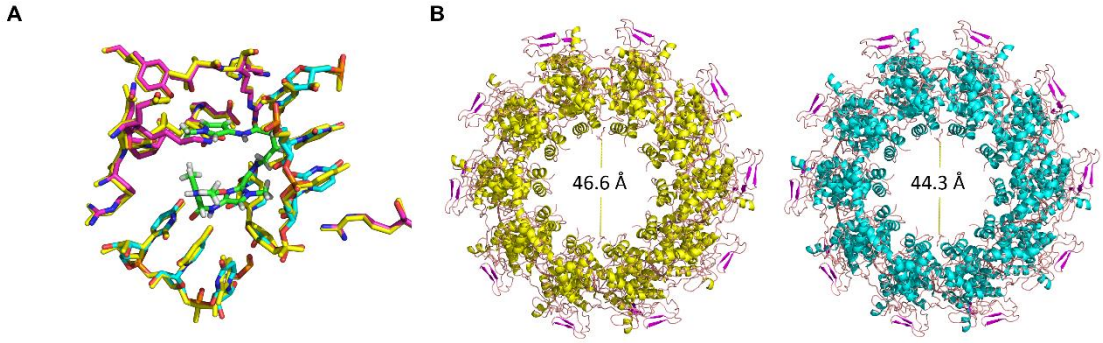


Figure 2.8 Global ring tightening of the NLP.

Global ring tightening of the NLP. (A) Overlay of 2GIC and UMSL1011 bound crystal structure. UMSL1011 is shown in green, while like previously, the magenta and cyan colors represent the protein and RNA, respectively. The 2GIC structure is colored yellow. There is little to no structural differences in the local area where UMSL1011 binds. (B) A reconstruction of the NLP ring. On the left-hand side is 2GIC (in yellow) and on the right-hand side is UMSL1011 structure (in cyan). An overall, global tightening of the ring by 2.3 Å is observed when measuring the inner diameter from E-chain to E-chain, corresponding to the largest change in the ring diameter.

Table 2.3 UMSL1011 and RNA interactions based on Nucleotide Sequence

	Carbonyl-3	Amide (N)-3	Im-4 (methyl)	Carbonyl-4	Amide (N)-4	γ-5	Carbonyl-6	Amide (N)-6
5							**A**,C	**A**
6								
7			Ribose-P					
8			Ribose-P		A,G,**C**			
9				**U**,A,C	**U**,G			
1								
2								
3		U,A,**G**						
4	U,**G**	**G**						

In the UMSL1011-NLP structure, each base was modeled (A, G, U, and C) at each nucleotide position and all possible hydrogen bonds were examined. The nucleotide position is numbered for each row, and each moiety of UMSL1011 is for each column. The interactions of each moiety of UMSL1011with the respective base or the backbone are placed in each grid. The larger letters indicate possible selectivity for each nucleotide derived from optimal interactions, and the box highlighted in orange indicates interaction with the sugar/phosphate backbone.

3 MUTATIONS IN THE NUCLEOCAPSID PROTEIN WERE COMPLEMENTED BY MUTATIONS IN THE L-PROTEIN TO RESTORE VIRAL RNA SYNTHESIS

3.1 Abstract

During viral RNA synthesis by the viral RNA-dependent RNA polymerase (vRdRp) of vesicular stomatitis virus, the sequestered RNA genome must be released from the nucleocapsid in order to serve as the template. Unveiling the sequestered RNA by interactions of vRdRp-proteins, the large subunit (L) and the phosphoprotein (P), with the nucleocapsid protein (N) must not disrupt the nucleocapsid assembly. We noticed that a flexible structural motif composed of an α-helix and a loop in the N-protein may act as the access gate to the sequestered RNA. This suggests that local conformational changes in this structural motif may be induced by interactions with the polymerase to unveil the sequestered RNA, without disrupting the nucleocapsid assembly. Mutations of several residues in this structural motif, Glu169, Phe171 and Leu174, to Ala resulted in loss of viral RNA synthesis in a minigenome assay. After implementing these mutations in the viral genome, mutant viruses were recovered by reverse genetics and serial passages. Sequencing the genomes of the mutant viruses revealed that compensatory mutations in L, P and N were required to restore the viral viability. Corresponding mutations were introduced in L, P and N, and their complementarity to the N mutations was confirmed by the minigenome assay. Introduction of the corresponding mutations is also sufficient to rescue the mutant viruses. These results suggested that the interplay of the N structural motif with the L-protein may play a role in accessing the nucleotide template without disrupting the overall structure of the nucleocapsid.

3.2 Introduction

Negative strand RNA viruses (NSVs) include a number of important human pathogens, such as Ebola, avian influenza, and respiratory syncytial virus. Understanding the mechanism of their replication assists in the development of new therapies. One unique feature that separates NSVs from other viral pathogens is that their genomic RNAs are sequestered in the nucleocapsid during viral RNA synthesis. NSVs encode their own viral RNA-dependent RNA polymerase (vRdRp). During viral RNA synthesis by vRdRp, the assembled nucleocapsid serves as the template, not the released RNA genome (118, 119). The mechanism by which vRdRp gains access to the sequestered genomic RNA to copy complementary RNAs is an unsolved question.

Vesicular stomatitis virus (VSV) is a prototypic NSV. VSV belongs to the Rhabdovirus family and its genome encodes five viral proteins, nucleocapsid (N), phosphoprotein (P), matrix (M), glycoprotein (G) and the large protein (L). vRdRp of VSV is composed of the L and P-proteins, whereas the nucleocapsid, assembled by polymerization of the N-proteins, serves as the template. The crystal structure of a nucleocapsid-like particle (NLP) shows that the assembly of VSV nucleocapsid requires extensive interactions of a long N-terminal arm and a large loop in the C-terminal domain of the N-protein between four neighboring subunits (89, 174). The RNA is sequestered between the N-terminal and C-terminal domains that are formed mostly with α-helices. Some of the bases in the sequestered RNA face the interior of the N-protein such that they could not be copied without first being exposed. One possible way is to induce an open N conformation as observed in the structure of some RNA-free N subunits (219, 220). However, that will also require untangling the interactions between the N subunits, which does not seem feasible. Another possible way is to induce a local conformational change at a proposed access gate in the N-protein, which will not disrupt the overall structure of the nucleocapsid. This

proposed access gate is one of helices in the N-terminal domain, helix 5, which covers the sequestered RNA. If vRdRp can induce a conformational change of helix 5, the sequestered RNA would be exposed to serve as the nucleotide template. The structure of the L-protein has been solved (117), as well as the nucleocapsid binding domain of the P-protein bound to the nucleocapsid (141). The orientation of the P domain bound to the nucleocapsid suggest that nucleocapsid bound vRdRp faces the access gate of the sequestered RNA. In this position, vRdRp may open the helix-5 access gate to release the sequestered RNA.

In this report, we tested the requirement of L-N interplay for viral RNA synthesis. Mutagenesis studies confirmed that the structural motif consisting of the helix 5 and the following loop may play a role in supporting viral RNA synthesis. Since the helix 5-loop motif is at one of the most flexible regions in the N-protein, vRdRp could readily induce a conformation change in this structural motif to unveil the sequestered RNA when bound to the nucleocapsid. The local structural change induced by vRdRp will not disrupt the overall structure of the nucleocapsid.

3.3 Results

3.3.1 The flexible structural motif in the N-protein

When the N and P-proteins of VSV were expressed in E. coli, a nucleocapsid-like particle (NLP) could be purified (144). The crystal structure of the NLP revealed that the N-protein has an extended N-terminal arm and a large C-terminal loop (89). The nucleocapsid is assembled by extensive interactions between the N subunits to sequester the viral RNA in the core (174). Once the nucleocapsid is assembled, the overall structure is fairly stable except for two regions as shown in **Figure 3.1**. Based on the B-factors, the C-terminal loop and the helix 5/subsequent loop (helix 5-loop) are the most flexible regions in the NLP. The structural motif composed of

the helix 5-loop forms the access gate to the sequestered RNA. In the NLP, the sequestered RNA is inaccessible if the helix 5-loop gate is in place. On the other hand, the flexible C-terminal loop is shown to interact with the nucleocapsid-binding domain of the P-protein (141). This interaction, however, still maintains the sequestering of the viral RNA. During the viral RNA synthesis, vRdRp must gain access to the sequestered RNA in order to use it as the nucleotide template. Since the helix 5-loop motif covers the sequestered RNA and is flexible, the sequestered RNA may be locally released if a conformational change of the helix 5-loop motif could be induced by vRdRp, which could be accomplished without interrupting the nucleocapsid assembly.

3.3.2 The role of the helix 5-loop motif

As shown in **Figure 3.1**, several residues are exposed on the surface of the helix 5-loop motif. Residues Arg153, Met157, Cys164, Lys165 and Asn168 are located on the helix 5, whereas Glu169, Phe171 and Leu174 are located on the loop. An alanine-scanning mutagenesis was carried out to verify the role of each residue in viral RNA synthesis. After each mutation was introduced in the plasmid expressing the N-protein, the activity of viral RNA synthesis was quantitated in a minigenome assay (221) (**Figure 3.2**). Mutations of most residues in helix 5 did not significantly change the activity, but obvious reduction was observed by the Met157Ala mutation. On the other hand, mutation of residues on the loop dramatically reduced the activity of viral RNA synthesis. This observation strongly suggests that the loop in this structural motif may play an important role in supporting viral RNA synthesis by vRdRp.

To ensure ubiquitously distributed protein concentrations, the protein expression level of N mutants was measured by western blot (**Figure 3.2C**). Most N mutants showed a similar level of protein expression except for L174A that had a 33% expression level compared to the wild

type N-protein. For most of the N mutants, the protein expression level was not the main factor responsible for reduction of viral RNA synthesis. In case of L174A, reduction in N-protein expression may not play a major role in reduction of viral RNA synthesis.

3.3.3 Nucleocapsid assembly of N mutants

To examine the effects of N mutations on the assembly of the nucleocapsid, these mutations were introduced in our expression system of a nucleocapsid-like particle (NLP) (144). The crystal structure of the NLP revealed the molecular interactions in the nucleocapsid (89) and NLP recapitulates the assembly of nucleocapsid as found in VSV virion (90). The N mutants were co-expressed with the P-protein in E. coli and NLPs were purified as in (144). The elution position of the N mutant complexes in the size exclusion chromatography confirmed that the N mutants formed the same NLP as the wild type N-protein (**Figure 3.3A**). Electron micrographs of negative stain images of the NLPs clearly showed the ring-like structure of mutant NLPs that contain 10 N subunits as the wild type NLP (**Figure 3.3B**).

3.3.4 Specific effects of N mutants on viral RNA synthesis

There are three types of viral RNA synthesized during VSV infection: mRNA during transcription, and cRNA/vRNA during replication. To dissect which specific types of viral RNA synthesis were affected by the N mutations, each species of the viral RNAs were quantitated by RT-qPCR in the minigenome assay (**Figure 3.4**). The minigenome contains the leader and trailer sequences of the full length genome and the luciferase reporter gene. As shown in **Figure 3.4**, the levels of mRNA and cRNA were reduced the most when N mutants were included in the minigenome assays. Since vRNA could also be produced by the minigenome plasmid, the level of vRNA in this assay would be the mixture of two sources. Based on the changes of vRNA levels, the level of mRNA was changed to a large degree compared to the wild type, consistent

with the reduction in luciferase activities. For L174A mutant that has a reduced level of protein expression, the vRNA level was similar to that when the wt N was included in the minigenome assay, suggesting that effects on viral RNA synthesis was not due to the protein level of the N mutant.

3.3.5 Compensatory mutations in other viral proteins

If the loop residues are involved in supporting viral RNA synthesis by vRdRp, it may be expected that compensatory mutations in other viral proteins can restore the activity of viral RNA synthesis by vRdRp. To identify the compensatory mutations, each of the mutations in the loop was introduced in the N gene of the genome of the parent VSV and mutant viruses were rescued by reverse genetics (222). At first, the mutant viruses were passaged five times in BHK-21 cells at 31 °C at a low MOI of 0.01 to avoid the potential formation of defective interfering (DI) particles. In the last round, mutant viruses were harvested 24 h post-infection, and virus titers were determined by plaque assays. Subsequently, mutant viruses passaged at 31 °C were used to infect BHK-21 cells at 37 °C with a MOI of 1.0, followed by three more passages using a MOI of 0.01 and titering by plaque assays. Serial passages were continued until the virus titers reached 1×107 PFU/mL (similar as wt VSV). The viruses from the last round were plaque purified in HeLa cells and the genome was sequenced for each mutant virus (**Table 3.1**). Compensatory mutations were identified in the L-protein of all three mutant viruses, confirming the critical role of the helix-loop motif in the interplay between the N and L-protein to support viral RNA synthesis. In addition, a compensatory mutation was also found in the P-protein in mutant Niii (Leu174Ala). This is also consistent with the role the helix-loop motif in supporting vRdRp activities because the interactions of P with N contribute to the activity of viral RNA synthesis (223, 224). More interestingly, mutations were also identified in the N-protein (**Table**

3.1). This suggests that optimal correlations between L and N are required for efficient viral growth. No mutations were found in M and G proteins.

We noticed that the mutant viruses passaged at 31 °C did not grow well when they were first transferred to 37 °C. This indicates that the initial compensatory mutations meet the minimum requirements to restore the viral RNA synthesis. Further compensatory mutations are required to fully restore the activity of vRdRp or effective virus assembly at 37 °C.

3.3.6 Verification of the compensatory mutations

Compensatory mutations identified in the rescued mutant viruses were introduced in the plasmids expressing N, P and L-proteins, respectively, and the minigenome assays were repeated. As shown in **Figure 3.5**, the compensatory mutation in the L-protein alone is not sufficient to restore the activity of viral RNA synthesis with the Ala mutations of Glu169 and Leu174 introduced in the N-protein. In case of Phe171, the mutation to Ala did not result in a large reduction of viral RNA synthesis and the single compensatory mutation of Arg998Gln in the L-protein is sufficient to restore the activity. Additional mutation of Leu110Val in the N-protein did not further enhance the activity of viral RNA synthesis, implying this mutation may be related only to the assembly of the virus. On the other hand, all compensatory mutations are required for Glu169 and Leu174 mutations in order to fully restore the activity. The minigenome assays confirmed that the Leu174Val mutation is required for full restoration of the polymerase activity.

3.3.7 Confirmation of compensatory mutations required for virus replication

To verify that the identified compensatory mutations are sufficient to construct a viable mutant virus, the mutations from the three mutant viruses were introduced in the genome of the parent VSV and virus rescue was carried out subsequently. Mutant viruses corresponding to all

N mutations were successfully rescued and grew to similar titers as the parent VSV and wt VSV (**Figure 3.6**). This result confirms that mutations identified in these mutant viruses can produce viable viruses when they are introduced in the parent virus.

3.4 Discussion

The mechanism of viral RNA synthesis by NSV vRdRp is unique because the template is the nucleocapsid, not the free RNA genome. Since the genomic RNA is sequestered in the nucleocapsid, vRdRp must induce structural changes in the N-protein in order to gain access to the sequestered RNA. Another important requirement is that the integrity of the nucleocapsid must be retained after the viral RNA synthesis is terminated. The size of vRdRp (117) suggests that the sequestered RNA may be locally released by vRdRp, without disrupting the overall structure of the nucleocapsid.

The atomic structure of the nucleocapsid and the N-protein has been determined for a number of NSVs, and in the form of NLP, empty NLP or the N subunit (85-87, 89, 147, 174, 203-205, 219, 220, 225-237). As shown by the vast amount of structural data, the common scheme of nucleocapsid assembly of NSVs involves extensive cross-subunit interactions by termini or loops extended from the core of the N-protein (112). These extensive interactions make the overall structure of the nucleocapsid very stable. This raises the question how vRdRp may gain access to the sequestered RNA. Based on an open conformation found in the RNA-free N subunit (219, 220), one possible mechanism is to induce the open conformation of the N-protein by vRdRp. However, this mechanism would require a large conformational change (up to 30 ° rotation) of the domains in the N-protein, and disruption of the cross-subunit interactions by the extended polypeptides. This not only requires broad interactions between vRdRP and the nucleocapsid, but also may result in disassembly of the nucleocapsid. The open conformation of

the N-protein is stabilized by the chaperone P-protein to serve as the precursor for nucleocapsid assembly. The assembly of the nucleocapsid may not be a reversible process after it has been completed and the P-protein been released from the assembled nucleocapsid.

In this study, we focused on a structural motif of the helix 5-loop in the N-protein of VSV that constitutes the access gate to the sequestered RNA. This structural motif is one of the two flexible regions in the nucleocapsid, suggesting its conformational changes are more easily induced by interactions with vRdRp. Structural changes of the helix 5-loop are expected to be local and would not disrupt the nucleocapsid assembly. Previous studies showed that residues on helix 5 directly interact with the sequestered genomic RNA and are essential for RNA encapsidation and viral RNA synthesis (89, 177). The structure of the nucleocapsid binding domain of the P-protein bound in the NLP indicated that vRdRp would face the helix 5-loop motif when it binds the nucleocapsid (141), implying a direct contact between them. Our mutagenesis studies proved that residues Met157, Glu169, Phe171 and Leu174 in the helix 5-loop motif are essential for the activity of viral RNA synthesis. The N mutations do not interrupt the assembly of the nucleocapsid and appear to mainly affect the synthesis of mRNA and cRNA. Compensatory mutations in vRdRp (L and P-proteins), as well as those in the N-protein, can restore the activity of viral RNA synthesis. These mutations are also sufficient to generate a viable mutant virus when introduced in the wt viral genome. Since the initial mutations we introduced in the N-protein were alanines, which eliminates the potential interactions of the wt N residues with other proteins, residues resulted from compensatory mutations may not directly re-establish interactions with the mutated Ala, but rather interactions with the helix 5-loop motif. This is also consistent with the observation that additional mutations in the N-protein are required to fully restore the activity of vRdRp. Nevertheless, our study confirmed that potential

interactions of the helix 5-loop motif with vRdRp may be required for viral RNA synthesis in VSV. A similar flexible structural motif is also found in the N-protein of paramyxoviruses (Nipha virus, parainfluenza virus 5 and mumps virus), suggesting a common mechanism for accessing sequestered RNA in NSV nucleocapsid (219, 233, 238).

3.5 Methods

3.5.1 Cell culture, viruses

BSR-T7 (a gift from Dr. Biao He, University of Georgia) and BHK-21 cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) containing 5% fetal bovine serum (FBS) with 100 units of penicillin, 20 units of streptomycin (221). BSR-T7 cells were additionally supplemented with Geneticin G418 (1mg/ml) on every second passage. HeLa cells were maintained in DMEM containing 10% FBS 100 units of penicillin, 20 units of streptomycin (239). Recombinant vaccinia virus vTF7-3 were prepared in CV-1 cells cultured in Eagle's Minimum Essential Medium (EMEM) supplemented with 10% FBS.

3.5.2 Plasmid mutagenesis

The original plasmids pLuc, pN, pP and pL were kindly provided by Dr. Asit K. Pattnaik at University of Nebraska-Lincoln. pLuc was constructed in pGEM-3 vector by inserting a leader sequence, the luciferase gene and a trailer sequence (194). pN, pP and pL were used to express N, P and L-proteins of VSV. All site- directed mutations in N, P and L genes were constructed using Q5 Site-Directed Mutagenesis Kit (New England Biolabs) following the manufacturer's instructions. The same method was also used to introduce site-directed mutations in the full length VSV genome.

The wt full-length VSV genome plasmid is a gift from Dr. Gail Wertz at the University of Alabama at Birmingham. The full-length VSV (Indiana serotype) genome was modified by

inserting the coding sequence of a mCherry protein fused at the N-terminus of the P gene (mcP VSV), using the overlap PCR method (**Table 3.2**).

3.5.3 VSV minigenome assay

BSR-T7 cells seeded to approximately 95% confluency in 12-well plates were infected with vTF7-3 at a MOI of 1-5 for 1 h at 37 °C. Cells were washed twice with Opti-MEM serum-free medium (Invitrogen) and transfected with 2µg pLuc, 1.5 µg of pN, 1 µg pP and 0.5 µg of pL, by using Lipofectamine 2000 (Invitrogen) (194). The cells were incubated for 6 h at 37 °C, and then fresh DMEM with 5% FBS was exchanged. At 24 hpt, cell lysates were assayed for the luciferase activity.

3.5.4 Western blot

Western blots were performed as described previously (240). Lysates were prepared from BSR-T7 cells transfected with N-WT, N-E169A, N-F171A and N-L174A. Samples were run on 8.5% acrylamide SDS-PAGE gels and Western blot was probed with specific antibodies for anti-VSV N and developed using PierceTM ECL Western Blotting Substrate (ThermoFischer Scientific).

3.5.5 N-protein expressions and purifications

All of cDNA mutations were cloned into the expression vector previously shown (117) using the HiFi assembly kit with the forward primer (5'-CTTAAGAAGGAGATATACCATGGCTTCTGTTACAGTCAAG-3') and the reverse primer (5'-GATGGCCCATGGTATATCTCCTTCATTGTCAAATTCTGAC-3') according to the manufacturer's protocol. Positive clones were identified by sequencing and transformed into E. coli strain BL21 (DE3). The N-rings were expressed and purified as previously described (144). After purification, size exclusion was carried out with a HiLoad 16/100 Superdex 200 prep grade

column equilibrated with 50 mM Tris pH 7.5 and 300 mM NaCl. The second peak, corresponding to the pure NLPS, was analyzed by SDS-page for purity, and then utilized for further TEM experiments.

3.5.6 Transmission Electron Microscopy

Images were taken at 30K magnification on a JEOL JEM-1400 (120 kV) TEM using a 2k x 2k, 14μm Ultrascan1000 (Gatan). Samples grids (25 μg/mL concentration of protein) were prepared by adding 2μL onto a carbon coated, 400 mesh copper grid. After 1 minute, the sample was blotted with Whatman 4 filter paper. A 2μL buffer wash/blot step was added before staining with 4% uranyl acetate for 30 seconds and blotting away excess. All particle picking and 2D class averages were done in EMAN2 and selected class averages and particles are shown in **Figure 3.3** (241).

3.5.7 Virus rescue

mcP VSV was used as the parent VSV to construct the mutant virus genome, following the procedures described previously (222, 242-244). Briefly, confluent 60-mm dishes of BSR-T7 cells were infected with vTF7-3 at a MOI of 10. After 45 min, the cells were co-transfected with a plasmid encoding the full length VSV genome containing various mutations in the corresponding viral genes (4 μg), and pN (6 μg), pP (4 μg) and pL (2 μg) proteins in 20 μL of Lipofectamine 2000. After 48-72 h, the culture medium was collected, filtered through a 0.22 μm-pore-size syringe filter twice to remove vTF7-3 vaccinia virus. Viral titers were determined by a plaque assay as described previously (245).

3.5.8 Plaque purification of rescued viruses.

Briefly, monolayers of Hela cells were grown in 12-wells tissue culture plates. Cells were infected with inoculums of the mutant virus 10-fold serially diluted in DMEM without serum.

After 1 h incubation at 37 °C in 5% CO2, the infection medium was then aspirated and washed three times with phosphate-buffered saline (PBS), and overlaid with 1 mL of 0.8% (wt/vol) agarose in DMEM. After 24 h incubation, four plaques were picked from at least 2 different wells. The agarose plugs were individually transferred to 60mm plates of BHK-21 cells. Virus growth was monitored by observing mCherry using fluorescence microscopy and virus titers were determined by plaque assays.

3.5.9 RNA extraction and reverse transcription-PCR (RT-PCR)

Viral RNAs were extracted from infected cell cultures by using the RNeasy Mini Kit (QIAGEN) following the manufacturer's instructions, and the total RNA was used in reverse transcription to generate cDNA by using random Nonamers (Sigma) as described previously (223, 246). Briefly, the 20 µL reaction mixtures contained 10 µL of RNA, 4 µL 5×first-strand buffer, 1 µL random nonamers primers (Sigma), 1 µL 10mM dNTP Mix, 2 µL 0.1 M DTT, 1 µL RNaseOUT inhibitor, and 1 µL of SuperScript II Reverse Transcriptase (Invitrogen). The samples were incubated at 42°C for 1 h, and the reaction was inactivated by heating at 70 °C for 15 min. The PCR reactions were performed by using Phusion High-Fidelity DNA Polymerase (Thermo Fisher).

3.5.10 Real-time qPCR

Total RNA was extracted, as described above, from BSR-T7 cells from either the mock, treated, or transfected with the different Luciferase minigenomes . The reverse transcriptions were performed using the same protocol as above, but by using the specific primers for (mRNA, 5'-CCAGATCGTTCGAGTCGTTTTTTTTTTTTTTTTTTTACACGGCGATCTTGCC-3'), cRNA, 5'-GCTAGCTTGAGCTAGGCATCCGCCGATATCTGTTAG-3'), vRNA, 5'-GGCCGTCATGGTGGCGAATACGAAGACAAACAAACC-3'), and (β-actin,

5’AGCACTGTGTTGGCGTACAG-3’). Real-time PCR was performed using the mRNA forward primer (5’-CCAGATCGTTCGAGTCGT-3’), and reverse primer (5’CGCGGTGGTGTTGTGTTC-3’), the cRNA forward primer (5’-GCTAGCTTCAGCTAGGCATC-3’) and reverse primer (5’-AAACAGAAAACCGACACCTG-3’), the vRNA forward primer (5’-AGCAGGTTTGTTGTACGC-3’) and reverse primer (5’GGCCGTCATGGTGGCGAAT-3’), and the β-actin forward primer (5’-AGAGCTACGAGCTGCCTGAC-3’) and reverse primer (5’-AGCACTGTGTTGGCGTAGAG-3’) on an QuantStudio 3 Real-Time PCR System utilizing the PowerUp Sybr Green Master Mix and the manufacturer’s protocol (Thermo Fisher Scientific). All data was analyzed using the 2-ΔΔCT method as previously described and the mock, the template free, the no RT-PCR, and the template only controls did not yield a CT value (data not included) after 40 cycles indicating specific amplification of the selected genes (247).

3.5.11 Sequence analysis of viral genomes

The primers shown in **Table 3.3** were used to amplify the full-length VSV genomes. Each fragment was cloned into pCRTM2.1 vector containing LacZ gene (Invitrogen). Automated sequence analysis was carried out by DNAStar 7.0 software. Determination of growth kinetics. Multi-step growth curves were generated using a protocol similar to that previously described (248). Briefly, BHK-21 cells seeded to approximately 95% confluency in 12-well plates were infected with each of the mutant viruses at a MOI of 0.01. After 1 h of adsorption at 37 °C in 5% CO2 with periodic rocking, the inoculum was removed. The cells were washed 3 times with PBS to remove unbound virus and then 1 mL of DMEM was added. At the indicated time points (6 h, 12h, 18 h, 24 h, 30 h, and 36 h), the supernatant of each well was collected and stored at -80 °C for determination of the virus titer by a plaque assay in a later time.

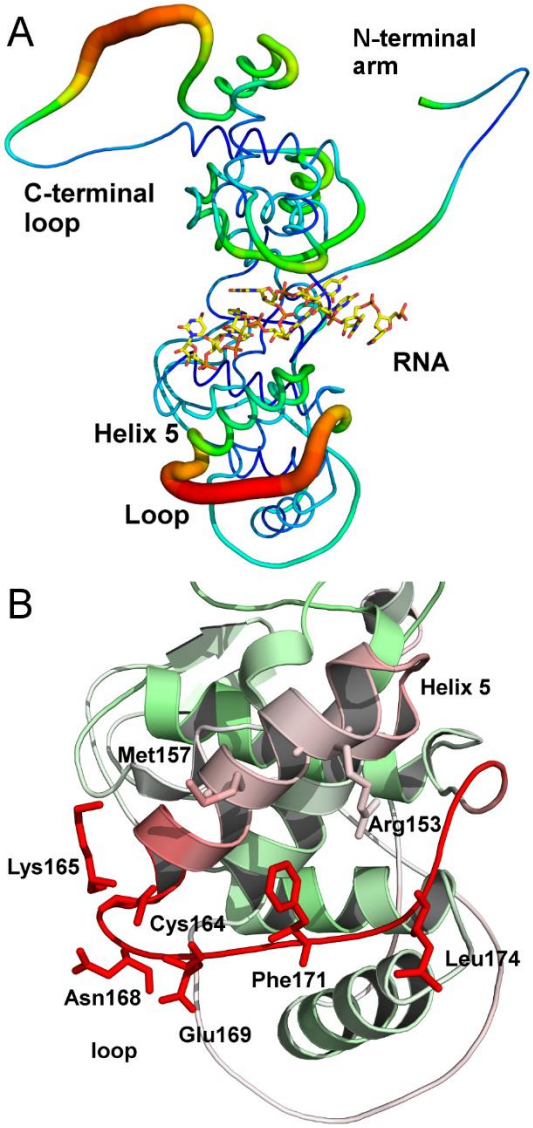


Figure 3.1 Flexibility of the N-protein

Flexibility of the N-protein. (A) A B-factor putty drawing of the N-protein structure in the nucleocapsid-like particle. The regions colored red also have a large diameter, which corresponds to higher B-factors and are associated with higher structural flexibility. The sequestered RNA was shown as a stick model. (B) A close-up look at the helix 5-loop motif. The red color of the models also corresponds to high B-factors. Selected residues are shown as stick models and labeled. The drawings were prepared with PyMol.

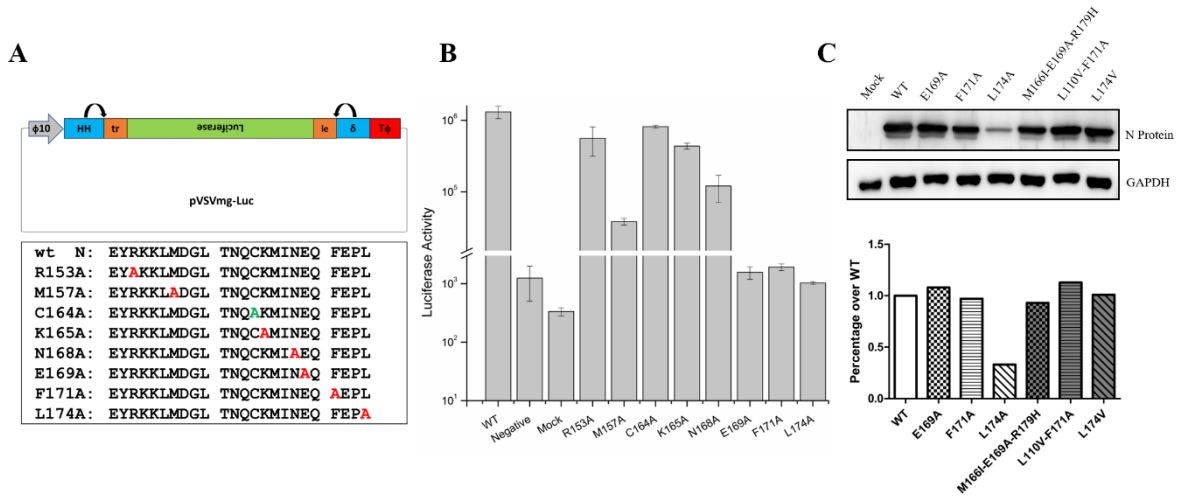


Figure 3.2 Role of each exposed residue.

Role of each exposed residue. (A) Illustration of the minigenome construct and the list of alanine mutations for the surface exposed residues in the helix 5-loop motif. Cys164 was also mutated to an alanine (green) as a control. (B) Luciferase activities in the minigenome assays. Mutations of the N-protein in pN and controls were labeled. (C) Protein expression levels of N mutants in the minigenome assays as determined by western blot.

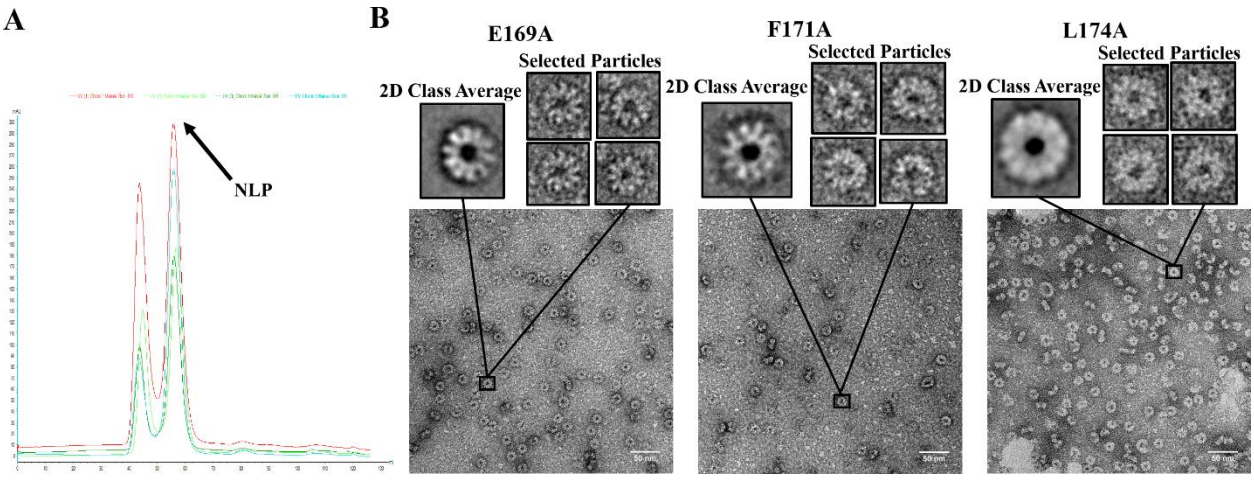


Figure 3.3 Mutations do not disrupt the NLP Structure

Mutations do not disrupt the NLP Structure. (A) Elution profile of NLPs from a S200 size exclusion column. The light blue line corresponds to wt NLP, the red line corresponds to E169A, the dark green line corresponds to F171A, and the light green line corresponds to L174A. (B) TEM images, class averages, and selected particles of the E169A, F171A, and L174A. Class averages were prepared with EMAN2 and particles were selected to show that the mutants still assemble in the same way as the wt rings (10 subunits in each assembly).

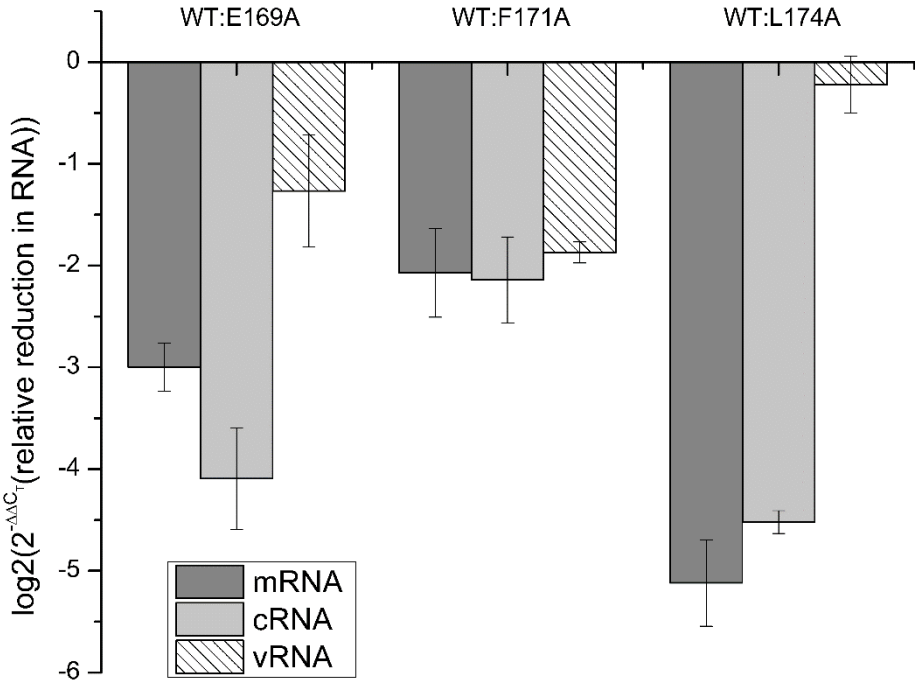


Figure 3.4 RT-qPCR quantification of mRNA, cRNA, and vRNA from the minigenome assay.

RT-qPCR quantification of mRNA, cRNA, and vRNA from the minigenome assay. Reduction in viral RNA levels of the E169A, F171A, and L174A mutants compared to the wt N, respectively, was calculated using $2^{-\Delta\Delta CT}$ values. The Log2 of $2^{-\Delta\Delta CT}$ is plotted in the vertical axis to show the relative reduction of RNA, and the shade for each viral RNA species is shown.

Table 3.1 Mutations identified in rescued viruses passaged at 37°C.

Virus	Passage	N gene	P gene	L gene
Ni (E169A)	14	M166I		I753M
		R179H		
Nii (F171A)	26	L110V		R998Q
Niii (L174A)	18	L174V	S264Y	R728K

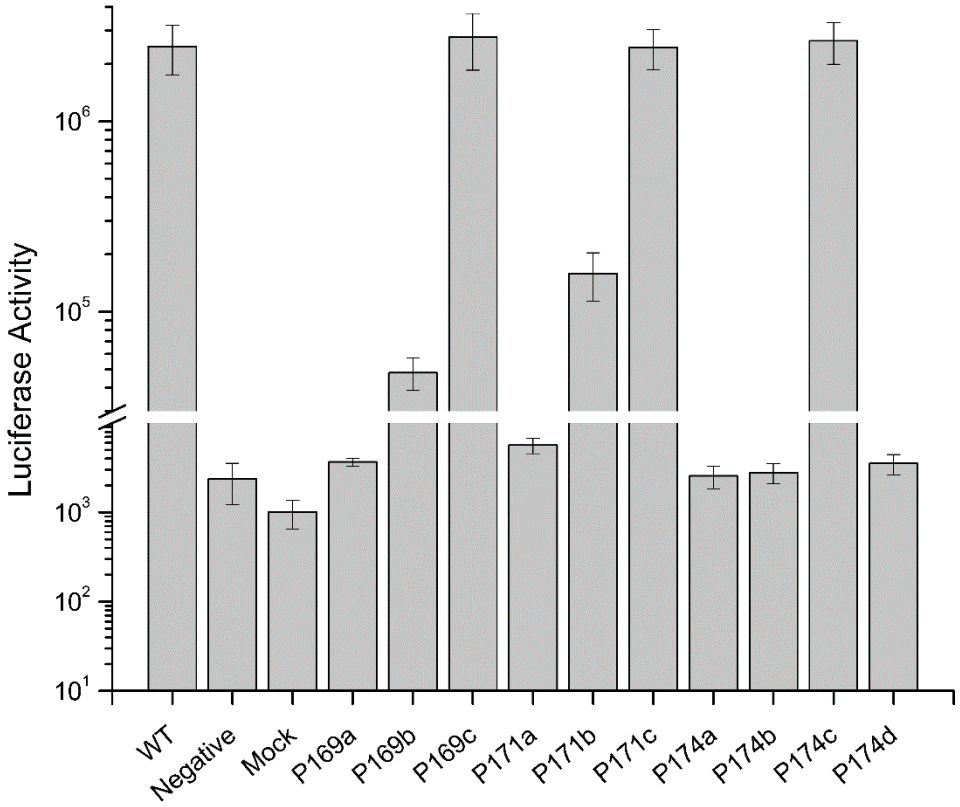


Figure 3.5 Verification of compensatory mutations.

Verification of compensatory mutations. Luciferase activities measured in the minigenome assays. The horizontal axis is labeled as the following: controls, P169a (pN-E169A, pP, and pL), P169b (pN-E169A, pP, pL-I753M), P169c (pN-M166I/E169A/R179H, pP, pL-I753M), P171a (pN-F171A, pP, pL), P171b (pN-F171A, pP, pL-R998Q), P171c (pN-L110V/F171A, pP, pL-R998A), P174a (pN-L174A, pP, pL), P174b (pN-L174A, pP, pL-R728K), P174c (pN-L174V, pP-S264Y, pL-R728K) and P174d (pN-L174A, pP-S264Y, pL-R728K). Mutations introduced in the plasmids that express the corresponding proteins or the wt plasmids are indicated in parentheses.

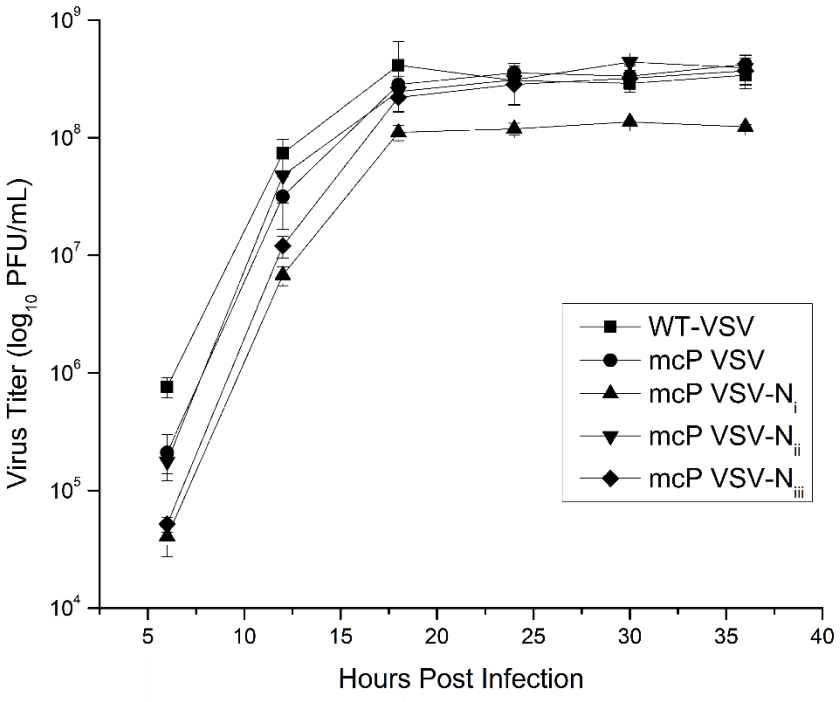


Figure 3.6 Growth curves of mutant viruses.

Growth curves of mutant viruses. Compensatory mutations identified in Ni, Nii and Niii were introduced in the genome of the parent VSV (mcP VSV) and the mutant viruses were rescued (mcP VSV-Ni, mcP VSV-Nii and mcP VSV-Niii, ***Table 3.1****). Confluent BHK-21 cells were infected by each marked virus with a MOI of 0.01. Samples were collected from each infection at the given time points and the virus titers were determined by plaque assays.*

Table 3.2 Primers used for introducing the mCherry gene.

Primer	Nucleotide sequence (5'—3')
N(+)-F	CAGCCTGATGACATTGAGTATACATCTCTTACTACAGCAGG
N-mC(-)-R	TCCTCGCCCTTGCTCACCAT*GGTGGCGGC*GATATCTGTTAGTTTTTTT
P(+)-F	ATGGACGAGCTGTACAAGggaggaaacagcggaggaATGGATAATCTCACAAAAGTTCG
P-M(-)-R	GCCTATTAAGTCATTATGCCAATTTAAATCTGAGCTTGACGGGC
mC(+)-F	TATGAAAAAAACTAACAGATATC*GCCGCCACC*ATGGTGAGCAAGGGCGAGG
mC-P(-)-R	ACTTTTGTGAGATTATCCATtcctccgctgtttcctccCTTGTACAGCTCGTCCAT

Table 3.3 Primers used for amplification of the full-length genome of VSV in sequencing.

Primer	Nucleotide sequence (5'—3')
AmpVSV1-F	ACGAAGACAAACAAACC
AmpVSV1-R	TCTGAAGTGCTCTGGTAC
AmpVSV2-F	TTCTAATCTAAAGGCCAC
AmpVSV2-R	CTTCAAGTTGCATATCGG
AmpVSV3-F	TAGGGAGGATGCAAACGG
AmpVSV3-R	TTGTCGGATAGGAGGTTG
AmpVSV4-F	CGGATTAATGAGAGCAAG
AmpVSV4-R	ACGAAGACCACAAAACCAG

4 CONSTRAINTS OF VIRAL RNA SYNTHESIS ON CODON USAGE OF NEGATIVE STRAND RNA VIRUS

At the time of submission under review at Journal of Virology, 11/18/2018

4.1 Abstract

Negative Strand RNA viruses (NSV) include some of the most pathogenic human viruses known. NSVs completely rely on the host cell for protein translation, but their codon usage bias is often different than that of the host. This discrepancy may have originated from the unique mechanism of NSV viral RNA synthesis in that the genomic RNA sequestered in the nucleocapsid serves as the template. The stability of the genomic RNA in the nucleocapsid appears to regulate its accessibility to the viral RNA polymerase, thus placing constraints on codon usage to balance viral RNA synthesis. By in situ analyses of vesicular stomatitis virus RNA synthesis, specific activities of viral RNA synthesis were correlated with the genomic RNA sequence. It was found that the purine/pyrimidine content modulates the overall stability of the polymerase complex, resulting in altering the activity of viral RNA synthesis. The codon usage is therefore constrained by the obligation of the NSV genome to viral RNA synthesis.

4.2 Introduction

The Negative strand RNA virus (NSV) family includes some of the most well-known human pathogens such as Ebola virus, influenza virus, and respiratory syncytial virus (77, 178-180). The virally encoded RNA-dependent RNA-polymerase (vRdRp) of NSV must recognize the nucleocapsid and unpack the sequestered genomic RNA from the nucleocapsid to use it as the template for viral RNA synthesis (118, 119). On the other hand, translation of viral proteins completely relies on the translation machinery of the host cell. This host reliance should direct

the evolution of the viral codon usage bias (CUB) towards that of the host. This is certainly an effective way to optimize virus growth if translation is the major barrier for the virus, as observed in positive strand RNA viruses (97).

However, there is no such coevolutionary virus-host relationship in the genome of NSVs. The genome of NSVs maintains a constant CUB, in spite of a high mutation rate by vRdRp (96). The calculated mutation rate for Paramyxoviridae is high enough to completely randomize the viral sequence 6 times per year, while the observed mutation rate is only about 6 nucleotide changes per year (96, 249-251). Furthermore, it has also been noted that the viral CUB can be significantly different from the host CUB (96). While this is often attributed to the suppression of CpG codons for evading the host immune system (252), the correlation is not ubiquitously distributed throughout all NSVs. Other requirements for virus growth may place constraints on evolution of NSVs. One factor could be the nucleotide content that is related to interactions of the genome with other proteins. Vesicular stomatitis virus (VSV) is a prototypic NSV that encodes five viral genes: nucleocapsid (N), phosphoprotein (P), matrix protein (M), glycoprotein (G) and large protein (L). A study showed that by altering the Codon Pair Bias score in a portion of the L-protein (polymerase) gene, which changes CUB, the virulence of VSV was attenuated in mice without changing the efficiency of viral protein translation (253).

Since NSV vRdRp needs to open the nucleocapsid to access the sequestered genomic RNA for transcription/replication, we propose that the stability of the genomic RNA in the nucleocapsid plays a regulatory role in the ability of the polymerase to successfully carry out viral RNA synthesis. As shown in **Figure 4.1A**, the nucleocapsid is formed through intricate cross-molecular interactions between adjacent subunits, and the accessibility of the genomic RNA could vary with the local sequence (89, 174). During viral RNA synthesis, one “stable”

sequence could cause the nucleocapsid to tighten and reduce the processivity of vRdRp, whereas a different "unstable" sequence could cause the nucleocapsid to loosen and increase the processivity of vRdRp or the rate of vRdRp dissociation from the nucleocapsid (**Figure 4.1B**). To verify this mechanism, in situ activities of VSV vRdRp were correlated with altered codon usage using minigenome assays. The results show that the balance between purines and pyrimidines in the genome sequence plays an essential role in regulating the polymerase activity. The requirements for transcriptional/replicational control constrain codon usage of NSVs, which explains why NSVs maintain their independent genomic stability despite reliance on the host machinery for viral protein translation.

4.3 Results

4.3.1 Effects of Changing the Codon Usage Bias on Transcription/Replication

In order to investigate the effect that CUB has on transcription/replication, CUB of the VSV genome was compared to that of human and the sand-fly, two of VSV hosts (**Table 4.4**). Furthermore, a similar analysis was extended to the genomes of Ebola virus and Influenza virus. Compared to CUB in human, codons highlighted in green are more frequently used in VSV and codons in red are less frequently used. It is obvious that CUB of VSV does not match that of human or sand-fly. A similar CUB pattern was also observed in Ebola virus and Influenza virus, indicating that this phenomenon could be ubiquitous throughout the NSV family.

A minigenome system was designed to quantitate in situ activities of vRdRp on VSV nucleocapsid (**Figure 4.2A**). The N gene was selected to represent the VSV genome. There are long stretches of green and short interspersed patches of red throughout the N gene (**Figure 4.2D**). This minigenome system is useful for comparable measurements of the vRdRp activities when the nucleotide content of the genome is altered. For comparisons with the wt minigenome

(247), one synthetic genome was recoded such that codons with low usage frequency (in red) were changed to those with high usage frequency (named as the high usage genome) and another synthetic genome was recoded such that codons with high usage frequency (in green) were changed to those with low usage frequency (named as the low usage genome). During the NSV life cycle, three species of RNA are produced: mRNA (during transcription), cRNA (during replication), and vRNA (during replication). In the minigenome assays, each RNA species was quantitated by RT-qPCR. In the top panel of **Figure 4.2B**, the full-length RNA levels from genomes with different CUBs were compared. For mRNA, ~744 fold reduction was observed for the high usage genome relative to the wt genome, and ~750 fold reduction, for the low usage genome. For cRNA, reduction of ~39 fold and ~55 fold were observed, respectively, and for vRNA, reduction of ~5.5 fold and ~1.8 fold were observed, respectively, for the high and low usage genomes. These observations clearly indicate that the activity of vRdRp is severely compromised during transcription and replication by just switching CUBs.

To further dissect which specific step in viral RNA synthesis is more affected by a change in CUB, eight chimeric minigenomes between the wt genome and either high or low usage genome were constructed (**Figure 4.2C**) based on the golden ratio. Among these chimeras, N1, N2, and N3 show statistically significant reduction in mRNA levels, while N4 shows a non-statistically significant increase. In the chimeric genome, N1, N2, and N3 all have either an altered CUB in the transcription (or replication) initiation or termination portion of the sequence, whereas N4 has the wt CUB in both. This would indicate that for transcription of mRNA, the CUB is more stringent for initiation and termination by vRdRp. This is consistent with the requirements for transcription by VSV vRdRp because viral transcription must be initiated and terminated at the correct position in order to balance the levels of all five mRNAs encoded in the

polycistronic genome. For cRNA and vRNA, on the other hand, it seems that elongation is the most affected activity of vRdRp since the initiation and termination of replication are more dependent on the 3' and 5' noncoding sequences (named the leader and trailer sequences, respectively), neither of which was changed in the chimeric genome. The largest reduction in genomic RNA levels was observed when the genomes with a large portion of altered CUB (N2 and N4) were used as the template. After initiation of replication, the processivity of vRdRp is dependent on CUB when it copies the entire genome.

It is possible that reduction in levels of full length RNAs was due to premature termination. Another set of primers was designed to quantify RNAs from any possible early termination during transcription and replication. The RNAs quantitated by these primers correspond to total RNAs of the shortened RNAs, in addition to the full length RNAs. As shown in **Figure 4.2B**, levels of the total mRNA transcripts were increased by ~19 fold and ~35 fold, respectively, for the high and low usage genomes. It is also interesting that the total mRNA was twice of the full length mRNA when the wt genome was used, indicating the common presence of early termination in transcription. An increase in shortened mRNA transcripts was observed across all of the chimeric minigenomes. This effect is exacerbated across the high usage chimeras, especially for N2, N3, and N4. On the other hand, an increase in shortened cRNA and vRNA was observed for the low usage chimeras, but less increase in shortened mRNA transcripts compared to those of the high usage chimeras.

4.3.2 The role of the Purine and Pyrimidine Ratio in Nucleocapsid Stability

In the nucleocapsid, the bases of the sequestered RNA stack in a consistent manner in each subunit (**Figure 4.1C**)(14). The first four bases stack with each other, the 5th base stacks with bases of 7 and 8, and the 6th base flips out of the stacked bases. The strength of π-stacking

interactions is in the order of purine-purine>purine-pyrimidine>pyrimidine-pyrimidine (254). We therefore hypothesized that sequences more rich in stacked purines would yield a more stable nucleocapsid, which is consistent with the structures of nucleocapsid-like particles (NLPs) that encapsidate poly(rA) (173, 255).

The sequence analysis technique Kappa Index of Coincidence (Kappa IC) was often used to identify different classes of promoters in eukaryotes (256). Kappa IC measures the probability of finding a repeating pattern in a nucleotide sequence. In this case, the purine content (A+G%) vs Kappa IC was plotted for the wt, high, and low usage genomes. Since Kappa IC is sensitive to simple or short tandem repeats, and window sizes, the calculation was carried out using various window sizes (**Figure 4.3A**) (256). Analysis of the template strand for mRNA synthesis revealed that there is a slightly lower degree of purine content for the high usage genome compared to that of the wt genome when the window size was 9 or 18 nucleotides. The more significant indication was that the lower (A+G%) in the high usage genome was associated with a higher KIC, shown by a large tail in the upper left portion of the graph. High KIC suggests that pyrimidines were more clustered in the high usage genome, leading to more pyrimidine stacking (257). As the window size increases, the left portion of the graph was more densely packed, further indicating the presence of larger stretches of pyrimidines. At the same time, the center of mass in each plot with smaller window sizes is consistently associated with higher Kappa IC scores for the high usage genome, consistent with a higher pyrimidine clustering at the local sequence level. When larger window sizes were used, the center of mass shifts slightly more to the right, almost overlapping with that of the wt genome (on the X-axis), indicating that the clustered pyrimidine content of the wt and high usage genomes matches more closely at a global level. This confirms that the local purine/pyrimidine clustering may play an important role in regulating vRdRp

activities in transcription. For the low usage genome, on the other hand, there is a significant tail to the right of the graph and right-shift of the center of mass throughout all window sizes. This is indicative of a higher purine content and clustering, while the overall clustering of purines and pyrimidines matches that of the wt genome.

To verify that the stability of the nucleocapsid is related to the purine/pyrimidine content, Thermal Shift Assays (TSA) were carried out with recombinantly purified NLPs (144, 173, 258). A selected piece of RNA could be reconstituted in the NLP by removing the original RNA with RNase A and subsequent incubation with the selected RNA (173). Melting curves of NLPs with poly(rA) and poly(rU) sequences were determined (**Figure 4**). The results showed that incorporation of poly(rA) in the NLP increased the stability by 2.33 ^{o}C compared to the recombinantly purified NLP with random RNA sequence (173), and 2.96 ^{o}C compared to the NLP with poly(rU). Furthermore, incorporation of poly(rA) in the NLP delayed the thermo-release of RNA by 7.86 ^{o}C compared to the recombinantly purified NLP with random RNA sequence, and 8.81 ^{o}C compared to the NLP with poly(rU) (258). Both of these experiments depict a large stabilization of the NLP by poly(rA). This is consistent with a previous study in which a polyamide (UMSL1011) that stabilizes the RNA in NLP delayed RNA thermo-release by 2.04 ^{o}C (258). These experimental results confirmed that the nucleocapsid stability is dependent on the purine content of sequestered RNA, which may change the accessibility of the genomic RNA in the nucleocapsid to vRdRp to modulate its transcription/replication activities.

4.4 Discussion

The high mutation rate of vRdRp has been related to virus evolvability and replication (259). In positive strand RNA viruses, the faster genome replication and optimal codon usage select viruses with good fitness (260-262). Since RNA viruses rely on the host cell for protein

translation, CUB of positive strand RNA viruses is similar to that of the host cell since its genomic RNA is also used as the messenger RNA for protein translation (263, 264). In NSVs, however, their CUB does not match that of the host cell despite their reliance on host protein translation. This phenomenon has previously been attributed to limiting CpG to evade the host's immune system and to a concept of codon constellation (96, 252, 264-266). More importantly, analyses of codon usages revealed that CUB of NSVs was not resulted from transitional selection (267-269). To identify other factors of CUB selection in NSVs, viral proteins that have close interactions with the coding regions of the viral genome are considered. One of the unique mechanisms that control viral genome replication of NSVs is that the template used by vRdRp in viral RNA synthesis is the nucleocapsid. The entire viral RNA genome is completely sequestered by the N-protein (85, 89) and its access to vRdRp may regulate viral RNA synthesis in NSVs (117, 173).We hypothesize that the access to the genomic RNA sequestered in the nucleocapsid constrains codon usage by NSVs, which is a determinantal factor in CUB selection during NSV evolution.

These constraints were examined in this study by evaluating in situ activities of VSV vRdRp using a minigenome assay. Our data showed that vRdRp became highly dysregulated with large decreases in transcription and replication when the codon usage in the VSV minigenome was deoptimized (**Figure 4.2B**). Changes in viral RNA synthesis were not related to protein translation since all viral proteins were produced by other vectors. This clearly confirmed that the nucleotide sequence of the viral genome in the nucleocapsid plays an essential role in regulating viral RNA synthesis by vRdRp. In an effort to identify which steps are more affected by changes of codon usage, eight chimeric minigenomes were also created. The results showed that the regions corresponding to initiation and termination of gene transcription had more

dramatic detrimental effects on vRdRp activities when CUB was deoptimized (**Figure 4.2B** and **4.2C**). In the context of the virus genome, this is consistent with that vRdRp is required to accurately initiate and terminate mRNA transcription in order to balance the transcript products from the polycistronic genome (270). For replication, elongation of cRNA and vRNA was most limited by CUB changes because the ends of the genome, which regulate the initiation and termination of genome replication by vRdRp, were not changed.

The consequence of CUB changes was analyzed for nucleotide contents and clustering. Since the crystal structures of VSV NLP suggest that purines have more stable base stacking (89, 173), an increase or decrease in (A+G%) and their clustering may alter the access of the sequestered genomic RNA to vRdRp. (A+G%) was calculated and compared with KIC for high and low usage genomes (**Figure 4.3A**). Since KIC is a measure of randomness, a deviation from the wt sequence could potentially deregulate viral RNA synthesis. The results showed that pyrimidine clustering was increased in the high usage genome and purine clustering was increased in the low usage genome. An abnormal stability of the nucleocapsid could not maintain the precise activities of vRdRp. To further corroborate the correlation of (A+G%) and their clustering with levels of viral RNA synthesis, thermal shift assays were used to measure the stability of the nucleocapsid when different RNA sequences were encapsidated. The results showed that NLPs were stabilized by the presence of purine sequences and destabilized by the presence of pyrimidine sequences in comparison with the recombinantly prepared NLP that encapsidates random RNAs. These data directly support that the nucleotide content and base clustering of the NSV genome, as a consequence of CUB selection, regulates the vRdRp accessibility of the genomic RNA sequestered in the nucleocapsid. The regional accessibility of the genomic RNA may be increased by reducing A+G content or clustering, but early

termination could also be increased. Evolution of NSVs results in their own CUB that balances mRNA transcripts and efficient genome replication.

Insights on the evolutionary principles of CUB in NSVs have great implications in managing outbreaks of highly pathogenic viruses, such as Ebola virus, influenza virus and rabies virus. Analyses of codon usage in these viruses (252, 269, 271, 272) found that these NSVs constantly mutate in different hosts, but their CUB was not coincident with that of the host. Our conclusion that optimal viral RNA synthesis regulated by the nucleocapsid is an additional determinant factor in CUB selection of NSVs will lead to an improved algorithm for mapping the evolution trajectory of these pathogens. That will greatly help in design of detection, vaccines and antiviral treatment for emerging NSVs.

4.5 Methods

4.5.1 Minigenome Plasmid

The pN, pP, and pL plasmids were provided by Dr. Asit K. Pattnaik at the University of Nebraska-Lincoln. The template for the minigenome was constructed as previously shown (194). To insert the N gene into the minigenome template, the Luc gene was removed from the minigenome using BamHI. The N gene was amplified from pN using primers as shown in **Table 4.1**, with BamHI cleavage sites on each side. This fragment was then cloned into the minigenome vector. In order to create a PCR-tag for the mRNA transcribed from the minigenome, mutations corresponding to primers shown in **Table 4.1** were introduced using the Q5 Site-Directed Mutagenesis Kit (New England Biolabs) following the manufacturers protocol. Furthermore, 3 stop codons were added after the start codon to prevent translation utilizing primers as shown in **Table 4.1** using the QuickChange II Mutagenesis Kit (Agilent Technologies) following the manufacturers protocol. High codon usage, Low codon usage, and

chimeric minigenomes were synthesized by Genscript (Piscataway, NJ) and cloned into the minigenome vector using the BamHI restriction enzyme sites.

4.5.2 Minigenome Assay

BSR-T7 cells (a gift from DR. Biao He, University of Georgia) were maintained in Dulbecco modified Eagle medium (DMEM) containing 5% fetal bovine serum (FBS) with 100 units of penicillin and 20 U of streptomycin (221). Plasmids were then transfected into the cells in 60 mm plates after cells reached approximately 95% confluency. Briefly, cells were washed twice with Opti-MEM serum-free media (Invitrogen) and then transfected with 2 μg of pN-minigenome, 1.5 μg of pN, 1 μg of pP, and 0.5 μg of pL using Lipofectamine 2000 (Invitrogen) (194). The cells were then incubated for 5 h at 37 °C and then fresh DMEM with 5% FBS was added. At 48 hpt, RNA was extracted from the cells using Trizol reagent (Ambion by Life Technologies) following the manufacturer's protocol. Immediately after RNA isolation 1 μg of RNA was used for RT-PCR using SuperScript II Reverse Transcriptase (Invitrogen) according to the manufacturer's protocol. The primers for RT-PCR were specific for mRNA, cRNA, and vRNA, as outlined in **Table 4.2**. Quantitative-PCR was performed on a QuantStudio 3 Real-Time PCR System using the primers in **Table 4.3** and the PowerUp Sybr Green Master Mix following the manufacturer's protocol (Thermo Fisher Scientific). The data were analyzed by the 2-ΔΔCT method as described previously (247). Furthermore, mock, template free, no RT-PCR, and the template only controls did not yield a CT value (data not included) after 40 cycles, indicating specific amplification of the selected genes (247). In addition, melting analysis revealed a single moiety for mRNA, cRNA, and vRNA for every minigenome vector.

4.5.3 Sequence Analysis

Codon usage frequencies of the hosts (**Table 4.4**) were downloaded from GenScript Codon Usage Frequency Table Tool, and calculated for VSV genome using Sequence Manipulation Suite in bioinformatics.org. Changes in usage fraction were defined as significant if the difference was $\geq+ \Delta 0.14$. Green indicates an increased fraction in NSV compared to human, whereas Red indicates a decreased fraction. Changes of nucleotides in VSV versus human codons are marked. For Lutzomyia longipalpis, fly codons are marked yellow if its fraction differs from that of VSV, or blue if it is the same as VSV ($\leq+\Delta 0.04$) but differs from human.

The Kappa Index of Coincidence was calculated as described previously, with the modification of plotting the X-axis as (A+G%) (256). All calculations were performed using Python (Python Software Foundation, version 3.5, https://www.python.org/) and the script is available at http://github.com/rgumpper/Biochemical_Tools . Data plots were drawn utilizing the open source plotting library for Python, Matplotlib (273). Sliding windows of 9, 18, 27, 36, 45, 54, 63, 72, 81, and 90 were used for KIC calculations to examine the difference between local and global genome structure. A window size divisible by 9 was used as the nucleocapsid encapsidates 9 nucleotides per N subunit.

4.5.4 Expression/Purification of NLP and Thermal Shift Assay (TSA)

The NLPs were expressed and purified as previously described (144). Briefly, after initial NiNTA purification size exclusion was carried out with a HiLoad 16/100 Superdex 200 prep grade column equilibrated with 50 mM Tris pH 7.5 and 300 mM NaCl. The NLP was then reconstituted with either poly(rA) or poly(rU) purchased from Midland (Midland, TX) as shown previously (173). Excess RNA was removed by further size exclusion purification and RNA

composition was checked by 260/280 ratio on a UV-vis spectrophotometer. NLPs containing RNA from recombinant purification, poly(rA), or poly(rU) were then subjected to TSA as previously described (258).

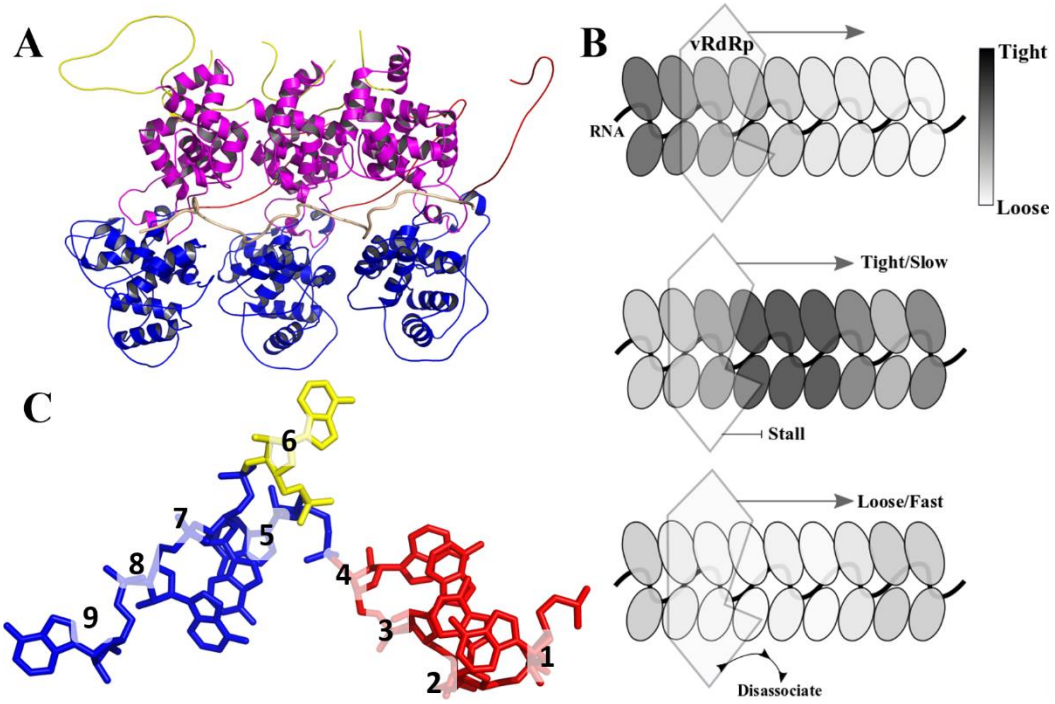


Figure 4.1 Structural interactions between the RNA and Nucleoprotein

Structural interactions between the RNA and Nucleoprotein. A. Ribbon representation of three Nucleocapsid protein subunits constructed from PDB 3PTX. Shown in red is the N-terminal arm, which interacts with adjacent subunits. Shown in blue and magenta are the N-lobe and the C-lobe respectively. Shown in yellow is the C-terminal loop, which also interacts with adjacent subunits. Furthermore, the backbone of the RNA is shown as a tan line sandwiched between the N and C-lobes. B. A cartoon representation of a tight or loose interaction of the genomic RNA in the nucleocapsid. This would regulate the accessibility of the sequestered RNA to vRdRp and cause possible disassociation or stalling of vRdRp. C. Stick representation of nine nucleotides (Adenosine) encapsidated in the nucleocapsid. Shown in red are the first 4 nucleotides that have bases stacked with each other. Shown in blue, are nucleotides 5, 7, and 8 that also have bases stacked. Nucleotide 6 is shown in yellow and does not directly interact with other nucleotides.

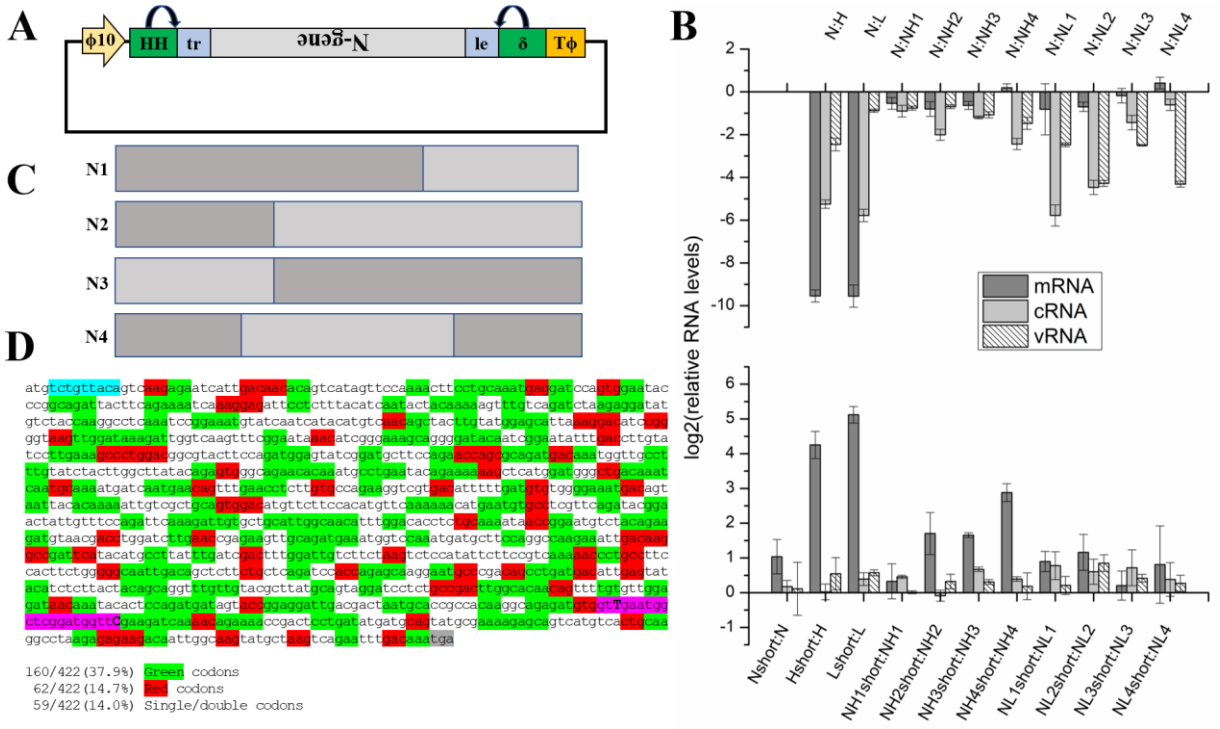


Figure 4.2 In-situ assay of vRdRp activity.

In-situ assay of vRdRp activity. A. Vector for the minigenome based on the N gene of VSV. Under the control of the T7 promoter, the vector will express the vRNA copy of the N gene, flanked by the leader and trailer sequences. B. Top panel shows the relative fold decreases in RNA levels from all the minigenomes quantitated by in situ vRdRp activity assays. N stands for the wt genome, H stands for the high usage genome and L stands for the low usage genome. Bottom panel shows the fold increases of shortened RNAs using primers that correspond to the beginning of each viral RNA moiety, as ratios to the full length RNA moieties. C. Construction of the chimeric minigenomes. Each was constructed based on the golden-ratio with the darker color representing the wt genome and lighter gray representing high or low usage genomes. Each chimeric genome is shown in panel B as NH1 or NL1 etc. D. Highlighted sequence of the N gene showing the location of the high frequency codons, in green, or the low frequency codons in red. Furthermore, highlighted in cyan is the placement of 3-stop codons to prevent translation of the mRNA transcripts. Also shown, in magenta, is the PCR-tag utilized in qPCR to identify mRNA transcripts transcribed only from the minigenome.

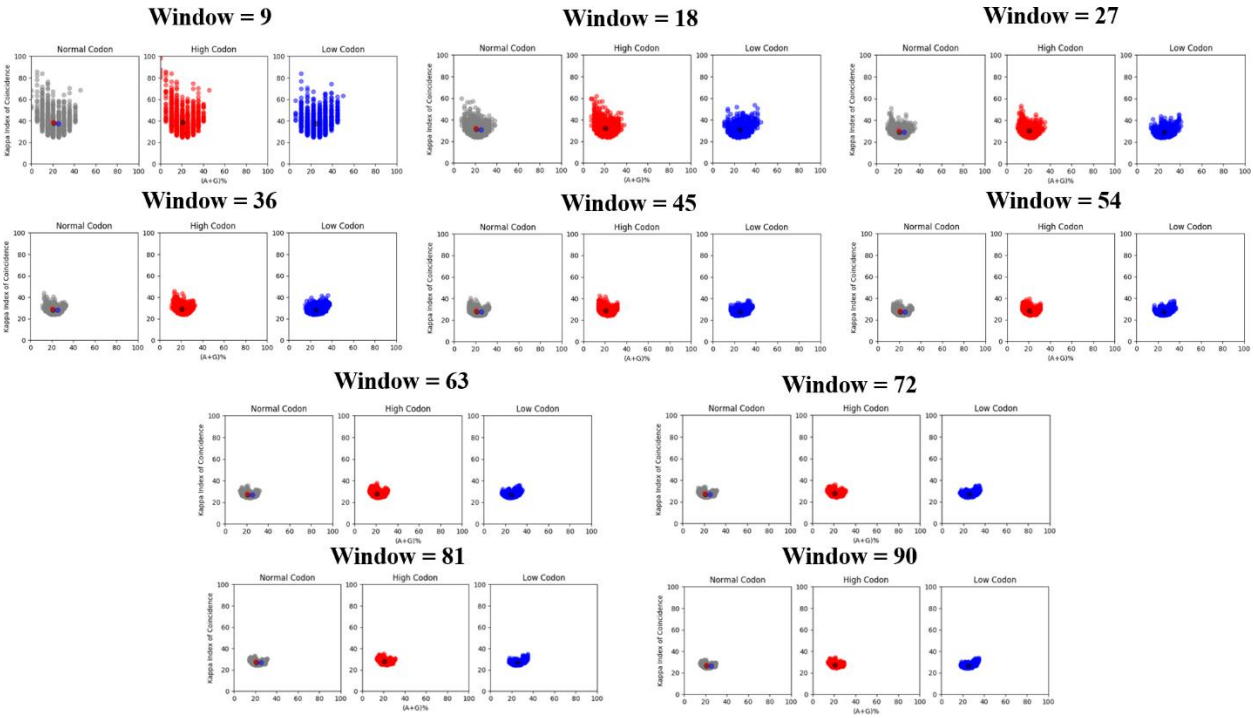


Figure 4.3 Calculated (A+G%) vs KIC.

Calculated (A+G%) vs KIC. Each panel is a different window size ranging from 9-90 as multiples of 9 due to the number of nucleotides in each N subunit. The wt genome is plotted in gray, the high usage genome is plotted in red, and the low usage genome is plotted in blue. The center of mass is plotted in their respective color for each panel, and all three centers of mass are plotted on the wt plot.

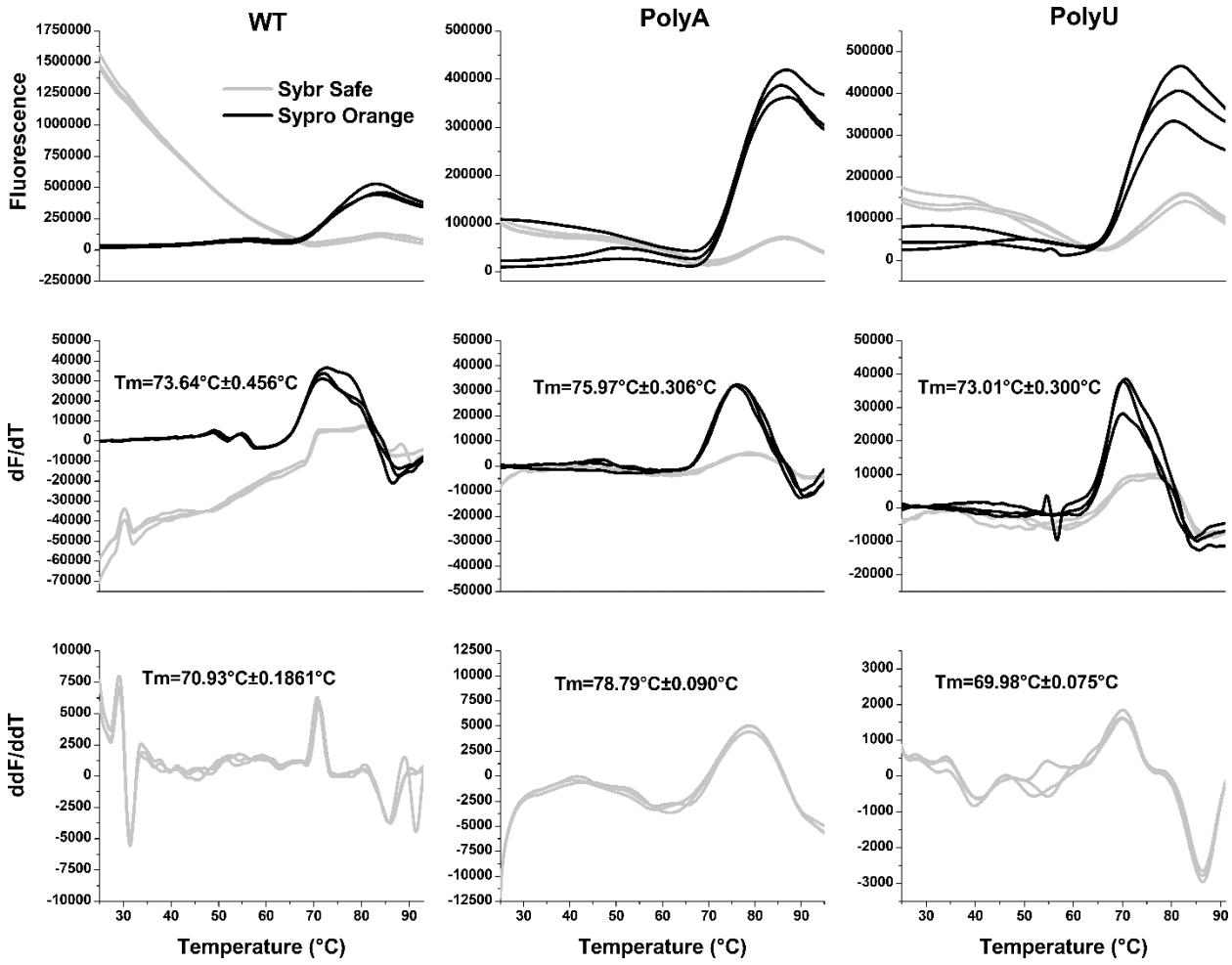


Figure 4.4 Thermal Shift Assays.

Thermal Shift Assays. Thermal shift assay was carried out on the randomly incorporated RNA in the recombinantly expressed NLP, poly(rA), and poly(rU) in reconstituted NLPs. Sybr Safe, which monitors RNA thermo-release from the NLP, is shown in gray, while Sypro Orange, which monitors protein denaturation is shown in black. Average Tm's with the standard deviation is shown on each panel.

Table 4.1 Primers for Minigenome Construction

	Minigenome Construction
N-gene-F	**5'-CGTACGATGTGATGATGAGTCAAGAG-3'**
N-gene-R	**5'-CGTACGTCATTTATCAAATTCTG-3'**
Triple Stop Codon-F	**5-'GTAATCACGTACGATGTGATGATGAGTCAAGAGAATCATTG-3'**
Triple Stop Codon-R	**5' CAATGATTCTCTTGACTCATCATCACATCGTACGTGATTAC-3'**
PCR-tag-F	**5'-CGGATGGTTCGAAGATCAAAACAGAAAACCG-3'**
PCR-tag-R	**5'-AGCCATTCAACCACATCTCTGCCTTGTGG-3'**

Table 4.2 RT-PCR primers

	RT-PCR primers
mRNA	**5'-CCAGATCGTTCGAGTCGTTTTTTTTTTTTTTTTCATTTGTCAAATTCTGACTTAG-3'**
cRNA	**5'-GCTAGCTTCAGCTAGGCATCCGCCGATATCTGTTAG-3'**
vRNA	**5'-GGCCGTCATGGTGGCGAATAGAAGTTTGGTAGGCTCG-3'**
mRNA-short	**5'-CCAGATCGTTCGAGTCGTTTGACATGTATGATTGATAC-3'**
cRNA-short	**5'-GCTAGCTTCAGCTAGGCATCCATCGTACGTGATTACTG-3'**
vRNA-short	**5'-GGCCGTCATGGTGGCGAATCGTACGTGATTACTGTTAAAG-3'**
β-Actin	**5'-AGCACTGTGTTGGCGTACAG-3'**

Table 4.3 Table of qPCR Primers

	qPCR Primers
mRNA-F	**5'-CCAGATCGTTCGAGTCGT -3'**
mRNA-R	**5'-AAGGCAGAGATGTGGTCG -3'**
cRNA-F	**5'-GCTAGCTTCAGCTAGGCATC-3'**
cRNA-R	**5'-AAACAGAAAACCGACTCCTG -3'**
vRNA-F	**5'-GGCCGTCATGGTGGCGAAT-3'**
vRNA-R	**5'-AGCAGGTTTGTTGTACGC -3'**
mRNA-s-F	**5'-CCAGATCGTTCGAGTCG-3'**
mRNA-s-R	**5'-ATGTGATGATGAGTC-3'**
cRNA-s-F	**5'-GCTAGCTTCAGCTAGGCATC-3'**
cRNA-s-R	**5'-ACGAAGACAAACAAACC-3'**
vRNA-s-F	**5'-GGCCGTCATGGTGGCGAAT-3'**
vRNA-s-R	**5'-ACGAAGACAAACAAAC-3'**
β-Actin-F	**5'-AGAGCTACGAGCTGCCTGAC-3'**
β-Actin-R	**5'-AGCACTGTGTTGGCGTACAG-3'**

Table 4.4 Fraction of Codon Usage

A.A.	Codon	Fraction				
		VSV	Human	Fly	EBOV	IFV
Ala	GCG	0.04	0.11	0.09	0.08	0.07
Ala	GCA	0.54	0.23	0.30	0.29	0.63
Ala	GCT	0.25	0.26	0.38	0.29	0.28
Ala	GCC	0.18	0.40	0.22	0.35	0.03
Cys	TGT	0.60	0.45	0.49	0.50	0.33
Cys	TGC	0.40	0.55	0.51	0.50	0.67
Asp	GAT	0.63	0.46	0.72	0.57	0.40
Asp	GAC	0.37	0.54	0.28	0.43	0.60
Glu	GAG	0.29	0.58	0.35	0.50	0.43
Glu	GAA	0.71	0.42	0.65	0.50	0.57
Phe	TTT	0.44	0.45	0.49	0.46	0.39
Phe	TTC	0.56	0.55	0.51	0.54	0.61
Gly	GGG	0.13	0.25	0.15	0.17	0.32
Gly	GGA	0.52	0.25	0.44	0.43	0.49
Gly	GGT	0.13	0.16	0.28	0.19	0.12
Gly	GGC	0.23	0.34	0.14	0.21	0.07
His	CAT	0.40	0.41	0.61	0.62	0.67
His	CAC	0.60	0.59	0.39	0.38	0.33
Ile	ATA	0.22	0.16	0.14	0.11	0.34
Ile	ATT	0.35	0.36	0.55	0.48	0.38
Ile	ATC	0.43	0.48	0.31	0.41	0.28
Lys	AAG	0.34	0.58	0.49	0.58	0.36
Lys	AAA	0.66	0.42	0.51	0.42	0.64
Leu	TTG	0.41	0.13	0.19	0.16	0.06
Leu	TTA	0.05	0.07	0.12	0.01	0.00
Leu	CTG	0.10	0.41	0.12	0.18	0.38
Leu	CTA	0.08	0.07	0.11	0.12	0.06
Leu	CTT	0.23	0.13	0.24	0.34	0.19
Leu	CTC	0.13	0.20	0.23	0.18	0.31
Met	ATG	1.00	1.00	1.00	1.00	1.00

		VSV	Human	Fly	EBOV	IFV
Asn	AAT	0.60	0.46	0.74	0.47	0.54
Asn	AAC	0.40	0.54	0.26	0.53	0.46
Pro	CCG	0.15	0.11	0.13	0.19	0.13
Pro	CCA	0.40	0.27	0.43	0.38	0.31
Pro	CCT	0.45	0.28	0.25	0.21	0.38
Pro	CCC	0.00	0.33	0.19	0.23	0.19
Gln	CAG	0.29	0.75	0.42	0.34	0.50
Gln	CAA	0.71	0.25	0.58	0.66	0.50
Arg	AGG	0.10	0.20	0.19	0.15	0.21
Arg	AGA	0.67	0.20	0.25	0.38	0.59
Arg	CGG	0.05	0.21	0.06	0.03	0.07
Arg	CGA	0.14	0.11	0.11	0.09	0.07
Arg	CGT	0.05	0.08	0.24	0.18	0.04
Arg	CGC	0.00	0.19	0.15	0.18	0.02
Ser	AGT	0.13	0.15	0.18	0.30	0.21
Ser	AGC	0.10	0.24	0.14	0.11	0.16
Ser	TCG	0.06	0.06	0.09	0.06	0.03
Ser	TCA	0.23	0.15	0.23	0.19	0.16
Ser	TCT	0.26	0.18	0.18	0.17	0.26
Ser	TCC	0.23	0.22	0.18	0.17	0.18
Thr	ACG	0.08	0.12	0.22	0.17	0.11
Thr	ACA	0.52	0.28	0.32	0.38	0.26
Thr	ACT	0.24	0.24	0.26	0.24	0.41
Thr	ACC	0.16	0.36	0.20	0.21	0.22
Val	GTG	0.23	0.47	0.23	0.19	0.48
Val	GTA	0.23	0.11	0.16	0.35	0.09
Val	GTT	0.19	0.18	0.44	0.26	0.22
Val	GTC	0.35	0.24	0.17	0.21	0.22
Trp	TGG	1.00	1.00	1.00	1.00	1.00
Tyr	TAT	0.43	0.43	0.44	0.57	0.58
Tyr	TAC	0.57	0.57	0.56	0.43	0.42
End	TGA	1.00				
End	TAG	0.00				
End	TAA	0.00				

5 SUMMARY

The goal of this work is to examine the basic structure-function relationships which guide viral replication via the nucleocapsid of VSV. Using a three-pronged approach towards understanding these basic principles, this work covers the discovery of a novel antiviral from an already widely understood class of compounds, the structure-based elucidation of a region important for viral polymerase assembly/processivity, and a paradigm shifting approach towards how the virus controls genetic drift of its genome.

Utilizing these techniques, a new potential antiviral compound was discovered that specifically targets the genomic RNA in the nucleoprotein. In doing so, it can specifically inhibit the activity of the vRdRp. This is one of the first compounds that can directly target the sequestered viral RNA encapsidated by a nucleocapsid. This compound was also found to be non-toxic to cells at higher concentrations than is needed to inhibit the activity of the virus. With further iterations, the efficacy of this polyamide will only continue to improve. Furthermore, these principles can be directly applied to explore if other polyamides can target other NSVs, more specifically, pandemic diseases like RSV. The drawback to this approach is that this polyamide may be specific for VSV, seeing as it recognizes both the genome and the protein. However, this study opens the door for further investigations and design of polyamides that can target the nucleocapsid of other more pathogenic viruses.

Our second approach, examines the fundamental interactions that guide the assembly and activity of the vRdRp. Through doing mutational studies on what we are calling the “access gate”, which is a helix positioned to protect the RNA, three key interactions were found. Three mutations E169A, F171A, and L174A were found to abolish the activity of the polymerase complex in a minigenome assay. Furthermore, incorporating these mutations into the

nucleoprotein does not disrupt the overall structure and assembly of the nucleocapsid. This would be an indicator that these interactions directly influence either the activity or the assembly of the vRdRp complex. Finally, after introducing these mutations into the viral genome and passaging the virus multiple times, we found compensatory mutations which restored the activity of the vRdRp. All the compensatory mutations that were found are on proteins associated with the polymerase complex. This indicates that these residues are crucial for the formation of an active vRdRp. With this information, we can probe the interactions that are imperative for the vRdRp, allowing us to conclude that the "access gate" helix is important for viral RNA transcription.

The last investigation reveals new insights into how NSVs control their genetic drift. Since RNA viruses are well known to have extremely high mutational rates, there must be constraints in place for these viruses to control genetic drift. Traditionally the main evolutionary bottleneck was thought to be controlled by the host's CUB, as in positive strand RNA viruses. However, this is not necessarily true when examining NSVs. This work has showed that the interactions between the genome and the nucleocapsid play a vital role in controlling the processivity of vRdRp. In turn, this would control genetic drift of the virus. Since the polymerase must access the RNA sequestered in the nucleocapsid, this would mean that the interactions between the genome and the nucleocapsid are central in controlling viral RNA transcription. The model states that there are some sections of the genome which will interact tightly with the nucleocapsid, making it harder for the polymerase to access and perhaps stalling/slowing it down. The inverse could also be true, loosely interacting sequences will make it too easy for polymerase to access the RNA and perhaps cause the complex to disassociate. Using an *in situ*

minigenome experiment, genomic sequence analysis, and stability assays, it was found that the pyrimidine/purine ratio plays a significant role in controlling the processivity of the vRdRp.

All this data garners interesting insights into how the nucleocapsid is a central role in the life cycle of VSV. Since the nucleoprotein encapsidates the genome, the structure-function properties that control its access are vital to the understanding of the viral replication life cycle. Furthermore, since these aspects are ubiquitous across NSVs these studies could be directly applied to more pathogenic viruses. As shown in this dissertation, understanding how this viral family controls and protects its genomic material is important not only for basic science, but also for the development of antivirals and vaccines. Hopefully this dissertation can shed light on some of these fundamental interests when studying NSV virology, and hopefully have a small impact on humanity at large.

5.1 Forward Looking Directions

Part of science is utilizing the data available to ask pertinent questions. In pushing the boundaries of known knowledge, it is important to be constantly innovating with a clear plan on how to move projects forward. In the context of the antiviral drug discovery portion of this dissertation, polyamides represent a promising class of compounds. UMSL1011 is the first compound which has been found to target the genomic RNA in an NSV. It represents an auspicious start towards using these compounds to target more pathogenic NSVs like RSV, Ebola, and even Influenza. Currently, UMSL1011 is specific for VSV and needs to be drastically improved. Verifying the nucleotide sequence that UMSL1011 targets would allow for the development of a set of rules which could be ubiquitously applied to other NSVs. Also, a targeted genomic approach can allow for functional studies which elucidate sensitive parts of the VSV genome. To make a more efficacious polyamide, without knowing the sequence that it

targets, one approach is to focus on a compound which interacts with Arg312 and Gln318. Getting a compound to bind in the pocket close to the RNA could disrupt the pertinent interactions involved in viral transcription. Another approach is to shorten the polyamide. The crystal structure does not contain the full-length polyamide. While it is impossible to tell whether this is due to an unstructured portion of the compound in the structure, or if the compound is getting hydrolyzed, shortening the polyamide may lead to a targeted effect. One could try the structured portion of the polyamide found in the crystal or design one with five linkers to target the RNA base stacking. This may lead to a stable compound able to target VSVs genome. Of course, the drawback would be an increase in off-target effects, as the promiscuity of the compound is going to increase as it gets shorter.

Studying the interactions which facilitate the assembly of the vRdRp has proven to be a challenging problem. The compensatory mutations identified in Chapter 3 lead to many interesting insights. Two of the mutations found on the L-protein are located on a single helix which borders the site where the template strand of RNA would enter for transcription to take place. One method to investigate if this region is important for producing an active vRdRp complex, is to identify all the solvent accessible residues on this helix and do alanine scanning mutations. Activity can then be tested by one of the *in situ* minigenome assays presented throughout this work. While this study would not allow one to identify if this helix is directly interacting with the nucleocapsid, it would build a stronger case for the access gate hypothesis. To directly test if this helix is interacting with the nucleocapsid, the construction of a fusion protein could be done. Replacing the N-terminal arm on the nucleocapsid with a flexible linker and the L-protein helix could allow one to recombinantly express a pseudo-vRdRp complex.

Structurally characterizing the fusion protein would be hard, but if *in situ* minigenome experiments show promising results, the likelihood of the fusion protein working would be high.

Being able to understand the constraints placed on the vRdRp and how it influences the mutation rates of NSVs is an interesting topic of study. While VSV is a great model system, replicating these results with an Ebola or RSV minigenome system would give credence to this ubiquitous method of evolutionary control throughout all NSVs. Furthermore, identifying the impact that this may have on different genes in VSV would be interesting. While it has already been shown that changing the codons in the L-protein can lead to viral attenuation, other genes have yet to be investigated in this manner. Further quantifying the amount of short terminating transcripts by the polymerase when altering the base stacking properties, is another direction that this study to go. Being able to tell where the polymerase would be stalling/dissociating would give a better idea of sequences that the polymerase is sensitive too. Finally, an interesting approach to understand how sequences control mutational constraints of the polymerase may lie in the fast-developing field of Deep Learning. Neural networks are now being extensively used in natural language processing. Nucleotide sequences are nothing but a language which we are not fluent in. Utilizing the large databases of viral sequences, which are tied to activity or attenuation of the virus and applying what we have found about the base stacking properties to these sequences, it is possible to build a deep learning model which could reveal unforeseen patterns and predict attenuating mutants.

While these are just some future directions that this work could take, sometimes the best scientific discoveries lie in the unpredictable. Most of the time an experiment does not tell you what you want, but it always tells you what is true. Following the truth will lead to interesting

discoveries and breakthrough scientific innovations. Nature always reveals herself in due time, it just takes a person who is willing to look and ask the right questions.

REFERENCES

1. **Knipe DM, Howley PM.** 2013. Fields virology, 6th ed. Wolters Kluwer/Lippincott Williams & Wilkins Health, Philadelphia, PA.
2. **Ivanofsky D.** 1903. On the mosaic disease of tobacco. Z Pfanzenkr **13:**1-41.
3. **Ivanofsky D.** 1892. Concerning the mosaic disease of the tobacco plant. St Petersburg Acad Imp Sci Bul **35:**67-70.
4. **Biejerink M.** 1898. Concerning a contagium vivum fluidum as a cause of the spot-disease of tobacco leaves. Verh Akad Wetensch **6:**3-21.
5. **d'Herelle F.** 1917. Sur un microbe invisible antagoniste des bacilles dysenteriques. C R Hebd Seances Acad Sci Paris **172:**99.
6. **Kaushce GA AP, Ruska H.** 1939. Die Sichtbarmachung von PF lanzlichem Virus in Ubermikroskop. Naturwissenschaften **27:**292-299.
7. **Loeffler F FP.** 1898. Zentralbl Bakteriol 1 Orig **28:**371.
8. **Reed W CJ, Agramonte A, Lazear J.** 1901. Senate Documents **66(822):**156.
9. **SS H.** 1977. The Virus: A History of the Concept. Heinemann Education Books, London.
10. **Nasir A, Kim KM, Caetano-Anolles G.** 2012. Viral evolution: Primordial cellular origins and late adaptation to parasitism. Mob Genet Elements **2:**247-252.
11. **Koonin EV, Senkevich TG, Dolja VV.** 2006. The ancient Virus World and evolution of cells. Biol Direct **1:**29.
12. **Koonin EV, Senkevich TG, Dolja VV.** 2009. Compelling reasons why viruses are relevant for the origin of cells. Nat Rev Microbiol **7:**615; author reply 615.
13. **Forterre P.** 2006. The origin of viruses and their possible roles in major evolutionary transitions. Virus Res **117:**5-16.
14. **Bandea CI.** 1983. A new theory on the origin and the nature of viruses. J Theor Biol **105:**591-602.
15. **La Scola B, Audic S, Robert C, Jungang L, de Lamballerie X, Drancourt M, Birtles R, Claverie JM, Raoult D.** 2003. A giant virus in amoebae. Science **299:**2033.
16. **Arslan D, Legendre M, Seltzer V, Abergel C, Claverie JM.** 2011. Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. Proc Natl Acad Sci U S A **108:**17486-17491.
17. **Philippe N, Legendre M, Doutre G, Coute Y, Poirot O, Lescot M, Arslan D, Seltzer V, Bertaux L, Bruley C, Garin J, Claverie JM, Abergel C.** 2013. Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. Science **341:**281-286.
18. **D'ellAmore C.** 2013. Biggest Virus Yet Foud, May Be Fourth Domain of Life?
19. **Moreira D, Lopez-Garcia P.** 2009. Ten reasons to exclude viruses from the tree of life. Nat Rev Microbiol **7:**306-311.
20. **Moreira D, Brochier-Armanet C.** 2008. Giant viruses, giant chimeras: the multiple evolutionary histories of Mimivirus genes. BMC Evol Biol **8:**12.
21. **Abroi A, Gough J.** 2011. Are viruses a source of new protein folds for organisms? - Virosphere structure space and evolution. Bioessays **33:**626-635.
22. **Forterre P.** 2005. The two ages of the RNA world, and the transition to the DNA world: a story of viruses and cells. Biochimie **87:**793-803.
23. **Forterre P.** 2013. The virocell concept and environmental microbiology. ISME J **7:**233-236.

24. **Margulis L.** 1998. Five Kingdoms: An Illustrated Guide to the Phyla of Life on Earth, 3 ed. WH Freeman, New York.
25. **Murphy FA FC, Bishop, DHL, et al.** 1995. Virus Taxonomy: The Classification and Nomenclature of Viruses. The Sixth Report of the International Committee on Taxonomy of Viruses. Springer-Verlag, Vienna.
26. **Baltimore D.** 1971. Expression of animal virus genomes. Bacteriol Rev **35:**235-241.
27. **Temin HM.** 1985. Reverse transcription in the eukaryotic genome: retroviruses, pararetroviruses, retrotransposons, and retrotranscripts. Mol Biol Evol **2:**455-468.
28. **Baltimore D.** 1985. Retroviruses and retrotransposons: the role of reverse transcription in shaping the eukaryotic genome. Cell **40:**481-482.
29. **Ryan KJ RC, eds.** 2004. Sherris Medical Microbiology. McGraw Hill.
30. **Mettenleiter ea.** 2008. Molecular Biology of Animal Herpesviruses. Canister Academic Press.
31. **Gubser C, Hue S, Kellam P, Smith GL.** 2004. Poxvirus genomes: a phylogenetic analysis. Journal of General Virology **85:**105-117.
32. **Fiers W SR.** 1962. The structure of the DNA of bacteriophage phiX174. Journal of Molecular Biology **5:**424.
33. **Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, Hutchison CA, Slocombe PM, Smith M.** 1977. Nucleotide sequence of bacteriophage phi X174 DNA. Nature **265:**687-695.
34. **Deng X, Yan Z, Cheng F, Engelhardt JF, Qiu J.** 2016. Replication of an Autonomous Human Parvovirus in Non-dividing Human Airway Epithelium Is Facilitated through the DNA Damage and Repair Pathways. PLoS Pathog **12:**e1005399.
35. **Wang XS, Yoder MC, Zhou SZ, Srivastava A.** 1995. Parvovirus B19 promoter at map unit 6 confers autonomous replication competence and erythroid specificity to adeno-associated virus 2 in primary human hematopoietic progenitor cells. Proc Natl Acad Sci U S A **92:**12416-12420.
36. **PH D.** 2015. Rotavirus Infection: A Disease of the Past? Infectious Disease Clinics of North America **29:**617-635.
37. **Tate JE, Burton AH, Boschi-Pinto C, Parashar UD, World Health Organization-Coordinated Global Rotavirus Surveillance N.** 2016. Global, Regional, and National Estimates of Rotavirus Mortality in Children <5 Years of Age, 2000-2013. Clin Infect Dis **62 Suppl 2:**S96-S105.
38. **Zell R, Delwart E, Gorbalenya AE, Hovi T, King AMQ, Knowles NJ, Lindberg AM, Pallansch MA, Palmenberg AC, Reuter G, Simmonds P, Skern T, Stanway G, Yamashita T, Ictv Report C.** 2017. ICTV Virus Taxonomy Profile: Picornaviridae. J Gen Virol **98:**2421-2422.
39. **Centers for Disease C.** July 23, 2018. HIV/AIDS Basic Statistics. https://www.cdc.gov/hiv/basics/statistics.html. Accessed
40. **Baltimore D.** 1970. RNA-dependent DNA polymerase in virions of RNA tumour viruses. Nature **226:**1209-1211.
41. **Temin HM, Mizutani S.** 1970. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. Nature **226:**1211-1213.
42. **Higgs MR, Chouteau P, Lerat H.** 2014. 'Liver let die': oxidative DNA damage and hepatotropic viruses. J Gen Virol **95:**991-1004.

43. **Lefkowitz EJ, Dempsey DM, Hendrickson RC, Orton RJ, Siddell SG, Smith DB.** 2018. Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). Nucleic Acids Res **46:**D708-D717.
44. **Davison AJ.** 2017. Journal of General Virology - Introduction to 'ICTV Virus Taxonomy Profiles'. J Gen Virol **98:**1.
45. **Anonymous.** Life Expectancy Data. https://www.kaggle.com/kumarajarshi/life-expectancy-who/version/1. Accessed
46. **Anonymous.** PDB Statistics: Overall Growth of Released Structures Per Year. https://www.rcsb.org/stats/growth/overall. Accessed
47. **Anonymous.** 02/02/2018. Novel Drug Approvals for 2017, *on* U.S. Food and Drug Administration. https://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugInnovation/ucm537040.htm. Accessed
48. **Baseler L, Chertow DS, Johnson KM, Feldmann H, Morens DM.** 2017. The Pathogenesis of Ebola Virus Disease. Annu Rev Pathol **12:**387-418.
49. **Molinari NA, Ortega-Sanchez IR, Messonnier ML, Thompson WW, Wortley PM, Weintraub E, Bridges CB.** 2007. The annual impact of seasonal influenza in the US: measuring disease burden and costs. Vaccine **25:**5086-5096.
50. **Erbelding EJ, Post DJ, Stemmy EJ, Roberts PC, Augustine AD, Ferguson S, Paules CI, Graham BS, Fauci AS.** 2018. A Universal Influenza Vaccine: The Strategic Plan for the National Institute of Allergy and Infectious Diseases. J Infect Dis **218:**347-354.
51. **Jedrzejas MJ, Singh S, Brouillette WJ, Laver WG, Air GM, Luo M.** 1995. Structures of aromatic inhibitors of influenza virus neuraminidase. Biochemistry **34:**3144-3151.
52. **White CL, Janakiraman MN, Laver WG, Philippon C, Vasella A, Air GM, Luo M.** 1995. A sialic acid-derived phosphonate analog inhibits different strains of influenza virus neuraminidase with different efficiencies. J Mol Biol **245:**623-634.
53. **Janakiraman MN, White CL, Laver WG, Air GM, Luo M.** 1994. Structure of influenza virus neuraminidase B/Lee/40 complexed with sialic acid and a dehydro analog at 1.8-A resolution: implications for the catalytic mechanism. Biochemistry **33:**8172-8179.
54. **Sugaya N, Tamura D, Yamazaki M, Ichikawa M, Kawakami C, Kawaoka Y, Mitamura K.** 2008. Comparison of the clinical effectiveness of oseltamivir and zanamivir against influenza virus infection in children. Clin Infect Dis **47:**339-345.
55. **Duval X, van der Werf S, Blanchon T, Mosnier A, Bouscambert-Duchamp M, Tibi A, Enouf V, Charlois-Ou C, Vincent C, Andreoletti L, Tubach F, Lina B, Mentre F, Leport C, Bivir Study G.** 2010. Efficacy of oseltamivir-zanamivir combination compared to each monotherapy for seasonal influenza: a randomized placebo-controlled trial. PLoS Med **7:**e1000362.
56. **Kawai N, Ikematsu H, Hirotsu N, Maeda T, Kawashima T, Tanaka O, Yamauchi S, Kawamura K, Matsuura S, Nishimura M, Iwaki N, Kashiwagi S.** 2009. Clinical effectiveness of oseltamivir and zanamivir for treatment of influenza A virus subtype H1N1 with the H274Y mutation: a Japanese, multicenter study of the 2007-2008 and 2008-2009 influenza seasons. Clin Infect Dis **49:**1828-1835.
57. **Cooper NJ, Sutton AJ, Abrams KR, Wailoo A, Turner D, Nicholson KG.** 2003. Effectiveness of neuraminidase inhibitors in treatment and prevention of influenza A and B: systematic review and meta-analyses of randomised controlled trials. BMJ **326:**1235.

58. **Mysore VS, Szablowski J, Dervan PB, Frost PJ.** 2016. A DNA-binding Molecule Targeting the Adaptive Hypoxic Response in Multiple Myeloma Has Potent Antitumor Activity. Mol Cancer Res **14:**253-266.
59. **Yamamoto M, Bando T, Morinaga H, Kawamoto Y, Hashiya K, Sugiyama H.** 2014. Sequence-specific DNA recognition by cyclic pyrrole-imidazole cysteine-derived polyamide dimers. Chemistry **20:**752-759.
60. **Kang JS, Meier JL, Dervan PB.** 2014. Design of sequence-specific DNA binding molecules for DNA methyltransferase inhibition. J Am Chem Soc **136:**3687-3694.
61. **Bremer RE, Szewczyk JW, Baird EE, Dervan PB.** 2000. Recognition of the DNA minor groove by pyrrole-imidazole polyamides: comparison of desmethyl- and N-methylpyrrole. Bioorg Med Chem **8:**1947-1955.
62. **White S, Baird EE, Dervan PB.** 1997. On the pairing rules for recognition in the minor groove of DNA by pyrrole-imidazole polyamides. Chem Biol **4:**569-578.
63. **Vasilieva E, Niederschulte J, Song Y, Harris GD, Jr., Koeller KJ, Liao P, Bashkin JK, Dupureur CM.** 2016. Interactions of two large antiviral polyamides with the long control region of HPV16. Biochimie **127:**103-114.
64. **He G, Vasilieva E, Harris GD, Jr., Koeller KJ, Bashkin JK, Dupureur CM.** 2014. Binding studies of a large antiviral polyamide to a natural HPV sequence. Biochimie **102:**83-91.
65. **Edwards TG, Koeller KJ, Slomczynska U, Fok K, Helmus M, Bashkin JK, Fisher C.** 2011. HPV episome levels are potently decreased by pyrrole-imidazole polyamides. Antiviral Res **91:**177-186.
66. **Fenner F NJ.** 1988. Poxviridae: The poxviruses, p 177-210, The laboratory Diagnosis of Infectious Diseases: Principles and Practice. Spring-Verlag.
67. **MR H.** 1992. Historical and contemporary perspectives in vaccine developments: From the vantage of cancer. , p 1-18, Progress in Medical Virology, vol 39. S Karger, Switzerland.
68. **Levine AJ.** 1994. The origins of the small DNA tumor viruses. Adv Cancer Res **65:**141-168.
69. **Fenner F.** 1980. The global eradication of smallpox. Med J Aust **1:**455-455.
70. **Henao-Restrepo AM, Camacho A, Longini IM, Watson CH, Edmunds WJ, Egger M, Carroll MW, Dean NE, Diatta I, Doumbia M, Draguez B, Duraffour S, Enwere G, Grais R, Gunther S, Gsell PS, Hossmann S, Watle SV, Konde MK, Keita S, Kone S, Kuisma E, Levine MM, Mandal S, Mauget T, Norheim G, Riveros X, Soumah A, Trelle S, Vicari AS, Rottingen JA, Kieny MP.** 2017. Efficacy and effectiveness of an rVSV-vectored vaccine in preventing Ebola virus disease: final results from the Guinea ring vaccination, open-label, cluster-randomised trial (Ebola Ca Suffit!). Lancet **389:**505-518.
71. **Henao-Restrepo AM, Longini IM, Egger M, Dean NE, Edmunds WJ, Camacho A, Carroll MW, Doumbia M, Draguez B, Duraffour S, Enwere G, Grais R, Gunther S, Hossmann S, Konde MK, Kone S, Kuisma E, Levine MM, Mandal S, Norheim G, Riveros X, Soumah A, Trelle S, Vicari AS, Watson CH, Keita S, Kieny MP, Rottingen JA.** 2015. Efficacy and effectiveness of an rVSV-vectored vaccine expressing Ebola surface glycoprotein: interim results from the Guinea ring vaccination cluster-randomised trial. Lancet **386:**857-866.

72. **Damm O, Witte J, Wetzka S, Prosser C, Braun S, Welte R, Greiner W.** 2016. Epidemiology and economic burden of measles, mumps, pertussis, and varicella in Germany: a systematic review. Int J Public Health **61:**847-860.
73. **Phadke VK, Bednarczyk RA, Salmon DA, Omer SB.** 2016. Association Between Vaccine Refusal and Vaccine-Preventable Diseases in the United States: A Review of Measles and Pertussis. JAMA **315:**1149-1158.
74. **Organization WH.** 2015. Health worker Ebola infections in Guinea, Liberia and Sierra Leone: A Preliminary Report. http://www.who.int/hrh/documents/21may2015_web_final.pdf. Accessed
75. **Paessler S, Walker DH.** 2013. Pathogenesis of the viral hemorrhagic fevers. Annu Rev Pathol **8:**411-440.
76. **Control TCfD.** 2018. Estimating Seasonal Influenza-Associated Deaths in the United States. https://www.cdc.gov/flu/about/disease/us_flu-related_deaths.htm. Accessed
77. **Bont L, Checchia PA, Fauroux B, Figueras-Aloy J, Manzoni P, Paes B, Simoes EA, Carbonell-Estrany X.** 2016. Defining the Epidemiology and Burden of Severe Respiratory Syncytial Virus Infection Among Infants and Children in Western Countries. Infect Dis Ther **5:**271-298.
78. **Gamarnik AV, Andino R.** 1998. Switch from translation to RNA replication in a positive-stranded RNA virus. Genes Dev **12:**2293-2304.
79. **Ortin J, Martin-Benito J.** 2015. The RNA synthesis machinery of negative-stranded RNA viruses. Virology **479-480:**532-544.
80. **Albertini AA, Ruigrok RW, Blondel D.** 2011. Rabies virus transcription and replication. Adv Virus Res **79:**1-22.
81. **Eisfeld AJ, Neumann G, Kawaoka Y.** 2015. At the centre: influenza A virus ribonucleoproteins. Nat Rev Microbiol **13:**28-41.
82. **Morin B, Kranzusch PJ, Rahmeh AA, Whelan SP.** 2013. The polymerase of negative-stranded RNA viruses. Curr Opin Virol **3:**103-110.
83. **Ruigrok RW, Crepin T, Kolakofsky D.** 2011. Nucleoproteins and nucleocapsids of negative-strand RNA viruses. Curr Opin Microbiol **14:**504-510.
84. **Reguera J, Cusack S, Kolakofsky D.** 2014. Segmented negative strand RNA virus nucleoprotein structure. Curr Opin Virol **5:**7-15.
85. **Albertini AA, Wernimont AK, Muziol T, Ravelli RB, Clapier CR, Schoehn G, Weissenhorn W, Ruigrok RW.** 2006. Crystal structure of the rabies virus nucleoprotein-RNA complex. Science **313:**360-363.
86. **Tawar RG, Duquerroy S, Vonrhein C, Varela PF, Damier-Piolle L, Castagne N, MacLellan K, Bedouelle H, Bricogne G, Bhella D, Eleouet JF, Rey FA.** 2009. Crystal structure of a nucleocapsid-like nucleoprotein-RNA complex of respiratory syncytial virus. Science **326:**1279-1283.
87. **Ye Q, Krug RM, Tao YJ.** 2006. The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA. Nature **444:**1078-1082.
88. **Cox R, Pickar A, Qiu S, Tsao J, Rodenburg C, Dokland T, Elson A, He B, Luo M.** 2014. Structural studies on the authentic mumps virus nucleocapsid showing uncoiling by the phosphoprotein. Proc Natl Acad Sci U S A **111:**15208-15213.
89. **Green TJ, Zhang X, Wertz GW, Luo M.** 2006. Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. Science **313:**357-360.

90. **Ge P, Tsao J, Schein S, Green TJ, Luo M, Zhou ZH.** 2010. Cryo-EM Model of the Bullet-Shaped Vesicular Stomatitis Virus. Science **327:**689-693.
91. **Whelan SPJ, Barr JN, Wertz GW.** 2004. Transcription and replication of nonsegmented negative-strand RNA viruses. Biology of Negative Strand Rna Viruses: The Power of Reverse Genetics **283:**61-119.
92. **Bean WJ, Jr., Cox NJ, Kendal AP.** 1980. Recombination of human influenza A viruses in nature. Nature **284:**638-640.
93. **Young JF, Palese P.** 1979. Evolution of human influenza A viruses in nature: recombination contributes to genetic variation of H1N1 strains. Proc Natl Acad Sci U S A **76:**6547-6551.
94. **Domingo E, Wain-Hobson S.** 2009. The 30th anniversary of quasispecies Meeting on 'Quasispecies: past, present and future'. Embo Reports **10:**444-448.
95. **Elena SF, Sanjuan R.** 2005. Adaptive value of high mutation rates of RNA viruses: separating causes from consequences. J Virol **79:**11555-11558.
96. **Rima BK.** 2015. Nucleotide sequence conservation in paramyxoviruses; the concept of codon constellation. J Gen Virol **96:**939-955.
97. **Coleman JR, Papamichail D, Skiena S, Futcher B, Wimmer E, Mueller S.** 2008. Virus attenuation by genome-scale changes in codon pair bias. Science **320:**1784-1787.
98. **Mueller S, Papamichail D, Coleman JR, Skiena S, Wimmer E.** 2006. Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. J Virol **80:**9687-9696.
99. **Martinez MA, Jordan-Paiz A, Franco S, Nevot M.** 2016. Synonymous Virus Genome Recoding as a Tool to Impact Viral Fitness. Trends Microbiol **24:**134-147.
100. **Werling D, Jungi TW.** 2003. TOLL-like receptors linking innate and adaptive immune response. Vet Immunol Immunopathol **91:**1-12.
101. **Carotti D, Palitti F, Santolamazza C, Lavia P, Strom R.** 1989. CpG frequency-dependence in the activity of eukaryotic DNA methyltransferase: in vitro methylation of CpG-rich islands. Ital J Biochem **38:**379A-382A.
102. **Bird AP.** 1980. DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res **8:**1499-1504.
103. **Chandler LA, Jones PA.** 1988. Hypomethylation of DNA in the regulation of gene expression. Dev Biol (N Y 1985) **5:**335-349.
104. **Simmonds P, Xia W, Baillie JK, McKinnon K.** 2013. Modelling mutational and selection pressures on dinucleotides in eukaryotic phyla--selection against CpG and UpA in cytoplasmically expressed RNA and in RNA viruses. BMC Genomics **14:**610.
105. **Rothberg PG, Wimmer E.** 1981. Mononucleotide and dinucleotide frequencies, and codon usage in poliovirion RNA. Nucleic Acids Res **9:**6221-6229.
106. **Krieg AM.** 2000. Immune effects and mechanisms of action of CpG motifs. Vaccine **19:**618-622.
107. **Sugiyama T, Gursel M, Takeshita F, Coban C, Conover J, Kaisho T, Akira S, Klinman DM, Ishii KJ.** 2005. CpG RNA: identification of novel single-stranded RNA that stimulates human CD14+CD11c+ monocytes. J Immunol **174:**2273-2279.
108. **Jensen S, Thomsen AR.** 2012. Sensing of RNA viruses: a review of innate immune receptors involved in recognizing RNA virus invasion. J Virol **86:**2900-2910.
109. **Tuller T, Zur H.** 2015. Multiple roles of the coding sequence 5' end in gene expression regulation. Nucleic Acids Res **43:**13-28.

110. **Faure G, Ogurtsov AY, Shabalina SA, Koonin EV.** 2016. Role of mRNA structure in the control of protein folding. Nucleic Acids Res **44:**10898-10911.
111. **Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW, Jr., Swanstrom R, Burch CL, Weeks KM.** 2009. Architecture and secondary structure of an entire HIV-1 RNA genome. Nature **460:**711-716.
112. **Green TJ, Cox R, Tsao J, Rowse M, Qiu S, Luo M.** 2014. Common mechanism for RNA encapsidation by negative-strand RNA viruses. J Virol **88:**3766-3775.
113. **Luo M, Green TJ, Zhang X, Tsao J, Qiu S.** 2007. Structural comparisons of the nucleoprotein from three negative strand RNA virus families. Virol J **4:**72.
114. **Hastie E, Cataldi M, Marriott I, Grdzelishvili VZ.** 2013. Understanding and altering cell tropism of vesicular stomatitis virus. Virus Res **176:**16-32.
115. **Obuchi M, Fernandez M, Barber GN.** 2003. Development of recombinant vesicular stomatitis viruses that exploit defects in host defense to augment specific oncolytic activity. J Virol **77:**8843-8856.
116. **Hastie E, Grdzelishvili VZ.** 2012. Vesicular stomatitis virus as a flexible platform for oncolytic virotherapy against cancer. J Gen Virol **93:**2529-2545.
117. **Liang B, Li Z, Jenni S, Rahmeh AA, Morin BM, Grant T, Grigorieff N, Harrison SC, Whelan SPJ.** 2015. Structure of the L-protein of Vesicular Stomatitis Virus from Electron Cryomicroscopy. Cell **162:**314-327.
118. **Emerson SU, Yu Y.** 1975. Both NS and L-proteins are required for in vitro RNA synthesis by vesicular stomatitis virus. J Virol **15:**1348-1356.
119. **Emerson SU, Wagner RR.** 1972. Dissociation and reconstitution of the transcriptase and template activities of vesicular stomatitis B and T virions. J Virol **10:**297-309.
120. **Kim IS, Jenni S, Stanifer ML, Roth E, Whelan SP, van Oijen AM, Harrison SC.** 2017. Mechanism of membrane fusion induced by vesicular stomatitis virus G protein. Proc Natl Acad Sci U S A **114:**E28-E36.
121. **Roche S, Bressanelli S, Rey FA, Gaudin Y.** 2006. Crystal structure of the low-pH form of the vesicular stomatitis virus glycoprotein G. Science **313:**187-191.
122. **Clague MJ, Schoch C, Zech L, Blumenthal R.** 1990. Gating kinetics of pH-activated membrane fusion of vesicular stomatitis virus with cells: stopped-flow measurements by dequenching of octadecylrhodamine fluorescence. Biochemistry **29:**1303-1308.
123. **Gaudin Y, Ruigrok RW, Knossow M, Flamand A.** 1993. Low-pH conformational changes of rabies virus glycoprotein and their role in membrane fusion. J Virol **67:**1365-1372.
124. **Durrer P, Gaudin Y, Ruigrok RW, Graf R, Brunner J.** 1995. Photolabeling identifies a putative fusion domain in the envelope glycoprotein of rabies and vesicular stomatitis viruses. J Biol Chem **270:**17575-17581.
125. **Roche S, Gaudin Y.** 2002. Characterization of the equilibrium between the native and fusion-inactive conformation of rabies virus glycoprotein indicates that the fusion complex is made of several trimers. Virology **297:**128-135.
126. **Nikolic J, Belot L, Raux H, Legrand P, Gaudin Y, A AA.** 2018. Structural basis for the recognition of LDL-receptor family members by VSV glycoprotein. Nat Commun **9:**1029.
127. **Ahmed M, McKenzie MO, Puckett S, Hojnacki M, Poliquin L, Lyles DS.** 2003. Ability of the matrix protein of vesicular stomatitis virus to suppress beta interferon gene

expression is genetically correlated with the inhibition of host RNA and protein synthesis. J Virol **77:**4646-4657.

128. **Gaudier M, Gaudin Y, Knossow M.** 2002. Crystal structure of vesicular stomatitis virus matrix protein. EMBO J **21:**2886-2892.
129. **Ahmed M, Lyles DS.** 1998. Effect of vesicular stomatitis virus matrix protein on transcription directed by host RNA polymerases I, II, and III. J Virol **72:**8413-8419.
130. **Raux H, Obiang L, Richard N, Harper F, Blondel D, Gaudin Y.** 2010. The matrix protein of vesicular stomatitis virus binds dynamin for efficient viral assembly. J Virol **84:**12609-12618.
131. **Henley JR, Cao H, McNiven MA.** 1999. Participation of dynamin in the biogenesis of cytoplasmic vesicles. FASEB J **13 Suppl 2:**S243-247.
132. **Rajani KR, Pettit Kneller EL, McKenzie MO, Horita DA, Chou JW, Lyles DS.** 2012. Complexes of vesicular stomatitis virus matrix protein with host Rae1 and Nup98 involved in inhibition of host transcription. PLoS Pathog **8:**e1002929.
133. **Petersen JM, Her LS, Varvel V, Lund E, Dahlberg JE.** 2000. The matrix protein of vesicular stomatitis virus inhibits nucleocytoplasmic transport when it is in the nucleus and associated with nuclear pore complexes. Mol Cell Biol **20:**8590-8601.
134. **Kieseier BC.** 2011. The mechanism of action of interferon-beta in relapsing multiple sclerosis. CNS Drugs **25:**491-502.
135. **Sitterlin D.** 2004. Characterization of the Drosophila Rae1 protein as a G1 phase regulator of the cell cycle. Gene **326:**107-116.
136. **Pritchard CE, Fornerod M, Kasper LH, van Deursen JM.** 1999. RAE1 is a shuttling mRNA export factor that binds to a GLEBS-like NUP98 motif at the nuclear pore complex through multiple domains. J Cell Biol **145:**237-254.
137. **Arai Y, Hosoda F, Kobayashi H, Arai K, Hayashi Y, Kamada N, Kaneko Y, Ohki M.** 1997. The inv(11)(p15q22) chromosome translocation of de novo and therapy-related myeloid malignancies results in fusion of the nucleoporin gene, NUP98, with the putative RNA helicase gene, DDX10. Blood **89:**3936-3944.
138. **Qanungo KR, Shaji D, Mathur M, Banerjee AK.** 2004. Two RNA polymerase complexes from vesicular stomatitis virus-infected cells that carry out transcription and replication of genome RNA. Proc Natl Acad Sci U S A **101:**5952-5957.
139. **Rahmeh AA, Schenk AD, Danek EI, Kranzusch PJ, Liang B, Walz T, Whelan SP.** 2010. Molecular architecture of the vesicular stomatitis virus RNA polymerase. Proc Natl Acad Sci U S A **107:**20075-20080.
140. **Morin B, Rahmeh AA, Whelan SP.** 2012. Mechanism of RNA synthesis initiation by the vesicular stomatitis virus polymerase. EMBO J **31:**1320-1329.
141. **Green TJ, Luo M.** 2009. Structure of the vesicular stomatitis virus nucleocapsid in complex with the nucleocapsid-binding domain of the small polymerase cofactor, P. Proc Natl Acad Sci U S A **106:**11713-11718.
142. **Schubert M, Harmison GG, Meier E.** 1984. Primary structure of the vesicular stomatitis virus polymerase (L) gene: evidence for a high frequency of mutations. J Virol **51:**505-514.
143. **Qiu S, Ogino M, Luo M, Ogino T, Green TJ.** 2016. Structure and Function of the N-Terminal Domain of the Vesicular Stomatitis Virus RNA Polymerase. J Virol **90:**715-724.

144. **Green TJ, Macpherson S, Qiu S, Lebowitz J, Wertz GW, Luo M.** 2000. Study of the assembly of vesicular stomatitis virus N protein: role of the P-protein. J Virol **74:**9515-9524.
145. **Ding H, Green TJ, Lu S, Luo M.** 2006. Crystal structure of the oligomerization domain of the phosphoprotein of vesicular stomatitis virus. J Virol **80:**2808-2814.
146. **Ding H, Green TJ, Luo M.** 2004. Crystallization and preliminary X-ray analysis of a proteinase-K-resistant domain within the phosphoprotein of vesicular stomatitis virus (Indiana). Acta Crystallogr D Biol Crystallogr **60:**2087-2090.
147. **Leyrat C, Yabukarski F, Tarbouriech N, Ribeiro EA, Jr., Jensen MR, Blackledge M, Ruigrok RW, Jamin M.** 2011. Structure of the vesicular stomatitis virus N(0)-P complex. PLoS Pathog **7:**e1002248.
148. **Masters PS, Banerjee AK.** 1988. Resolution of multiple complexes of phosphoprotein NS with nucleocapsid protein N of vesicular stomatitis virus. J Virol **62:**2651-2657.
149. **Peluso RW.** 1988. Kinetic, quantitative, and functional analysis of multiple forms of the vesicular stomatitis virus nucleocapsid protein in infected cells. J Virol **62:**2799-2807.
150. **Barik S, Banerjee AK.** 1991. Cloning and expression of the vesicular stomatitis virus phosphoprotein gene in Escherichia coli: analysis of phosphorylation status versus transcriptional activity. J Virol **65:**1719-1726.
151. **Barik S, Banerjee AK.** 1992. Phosphorylation by cellular casein kinase II is essential for transcriptional activity of vesicular stomatitis virus phosphoprotein P. Proc Natl Acad Sci U S A **89:**6570-6574.
152. **Bell JC, Prevec L.** 1985. Phosphorylation sites on phosphoprotein NS of vesicular stomatitis virus. J Virol **54:**697-702.
153. **Marnell LL, Summers DF.** 1984. Characterization of the phosphorylated small enzyme subunit, NS, of the vesicular stomatitis virus RNA polymerase. J Biol Chem **259:**13518-13524.
154. **Hsu CH, Kingsbury DW.** 1985. Constitutively phosphorylated residues in the NS protein of vesicular stomatitis virus. J Biol Chem **260:**8990-8995.
155. **Liu Z, Huntley CC, De BP, Das T, Banerjee AK, Oglesbee MJ.** 1997. Phosphorylation of canine distemper virus P-protein by protein kinase C-zeta and casein kinase II. Virology **232:**198-206.
156. **Jackson RL, Spadafora D, Perrault J.** 1995. Hierarchal constitutive phosphorylation of the vesicular stomatitis virus P-protein and lack of effect on P1 to P2 conversion. Virology **214:**189-197.
157. **Pattnaik AK, Hwang L, Li T, Englund N, Mathur M, Das T, Banerjee AK.** 1997. Phosphorylation within the amino-terminal acidic domain I of the phosphoprotein of vesicular stomatitis virus is required for transcription but not for replication. J Virol **71:**8167-8175.
158. **Hwang LN, Englund N, Das T, Banerjee AK, Pattnaik AK.** 1999. Optimal replication activity of vesicular stomatitis virus RNA polymerase requires phosphorylation of a residue(s) at carboxy-terminal domain II of its accessory subunit, phosphoprotein P. J Virol **73:**5613-5620.
159. **Mondal A, Victor KG, Pudupakam RS, Lyons CE, Wertz GW.** 2014. Newly identified phosphorylation site in the vesicular stomatitis virus P-protein is required for viral RNA synthesis. J Virol **88:**1461-1472.

160. **Emerson SU, Wagner RR.** 1973. L-protein requirement for in vitro RNA synthesis by vesicular stomatitis virus. J Virol **12:**1325-1335.
161. **Hunt DM, Smith EF, Buckley DW.** 1984. Aberrant polyadenylation by a vesicular stomatitis virus mutant is due to an altered L-protein. J Virol **52:**515-521.
162. **Ogino T, Banerjee AK.** 2007. Unconventional mechanism of mRNA capping by the RNA-dependent RNA polymerase of vesicular stomatitis virus. Mol Cell **25:**85-97.
163. **Tekes G, Rahmeh AA, Whelan SP.** 2011. A freeze frame view of vesicular stomatitis virus transcription defines a minimal length of RNA for 5' processing. PLoS Pathog **7:**e1002073.
164. **Jacome R, Becerra A, Ponce de Leon S, Lazcano A.** 2015. Structural Analysis of Monomeric RNA-Dependent Polymerases: Evolutionary and Therapeutic Implications. PLoS One **10:**e0139001.
165. **Pflug A, Guilligay D, Reich S, Cusack S.** 2014. Structure of influenza A polymerase bound to the viral RNA promoter. Nature **516:**355-360.
166. **Reich S, Guilligay D, Pflug A, Malet H, Berger I, Crepin T, Hart D, Lunardi T, Nanao M, Ruigrok RW, Cusack S.** 2014. Structural insight into cap-snatching and RNA synthesis by influenza polymerase. Nature **516:**361-366.
167. **Rahmeh AA, Li J, Kranzusch PJ, Whelan SP.** 2009. Ribose 2'-O methylation of the vesicular stomatitis virus mRNA cap precedes and facilitates subsequent guanine-N-7 methylation by the large polymerase protein. J Virol **83:**11043-11050.
168. **Egloff MP, Benarroch D, Selisko B, Romette JL, Canard B.** 2002. An RNA cap (nucleoside-2'-O-)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization. EMBO J **21:**2757-2768.
169. **Ray D, Shah A, Tilgner M, Guo Y, Zhao Y, Dong H, Deas TS, Zhou Y, Li H, Shi PY.** 2006. West Nile virus 5'-cap structure is formed by sequential guanine N-7 and ribose 2'-O methylations by nonstructuraL-protein 5. J Virol **80:**8362-8370.
170. **Zhou Y, Ray D, Zhao Y, Dong H, Ren S, Li Z, Guo Y, Bernard KA, Shi PY, Li H.** 2007. Structure and function of flavivirus NS5 methyltransferase. J Virol **81:**3891-3903.
171. **Melzer MK, Lopez-Martinez A, Altomonte J.** 2017. Oncolytic Vesicular Stomatitis Virus as a Viro-Immunotherapy: Defeating Cancer with a "Hammer" and "Anvil". Biomedicines **5**.
172. **Felt SA, Grdzelishvili VZ.** 2017. Recent advances in vesicular stomatitis virus-based oncolytic virotherapy: a 5-year update. J Gen Virol doi:10.1099/jgv.0.000980.
173. **Green TJ, Rowse M, Tsao J, Kang J, Ge P, Zhou ZH, Luo M.** 2011. Access to RNA encapsidated in the nucleocapsid of vesicular stomatitis virus. J Virol **85:**2714-2722.
174. **Zhang X, Green TJ, Tsao J, Qiu S, Luo M.** 2008. Role of intermolecular interactions of vesicular stomatitis virus nucleoprotein in RNA encapsidation. J Virol **82:**674-682.
175. **Bell JC, Brown EG, Takayesu D, Prevec L.** 1984. Protein kinase activity associated with immunoprecipitates of the vesicular stomatitis virus phosphoprotein NS. Virology **132:**229-238.
176. **Leyrat C, Jensen MR, Ribeiro EA, Jr., Gerard FC, Ruigrok RW, Blackledge M, Jamin M.** 2011. The N(0)-binding region of the vesicular stomatitis virus phosphoprotein is globally disordered but contains transient alpha-helices. Protein Sci **20:**542-556.

177. **Rainsford EW, Harouaka D, Wertz GW.** 2010. Importance of hydrogen bond contacts between the N protein and RNA genome of vesicular stomatitis virus in encapsidation and RNA synthesis. J Virol **84:**1741-1751.
178. **Appolinario CM, Jackson AC.** 2015. Antiviral therapy for human rabies. Antivir Ther **20:**1-10.
179. **Balmith M, Faya M, Soliman ME.** 2017. Ebola virus: A gap in drug design and discovery - experimental and computational perspective. Chem Biol Drug Des **89:**297-308.
180. **Wu W, Liu S.** 2017. The Drug Targets and Antiviral Molecules for Treatment of Ebola Virus Infection. Curr Top Med Chem **17:**361-370.
181. **Hiraoka K, Inoue T, Taylor RD, Watanabe T, Koshikawa N, Yoda H, Shinohara K, Takatori A, Sugimoto H, Maru Y, Denda T, Fujiwara K, Balmain A, Ozaki T, Bando T, Sugiyama H, Nagase H.** 2015. Inhibition of KRAS codon 12 mutants using a novel DNA-alkylating pyrrole-imidazole polyamide conjugate. Nat Commun **6:**6706.
182. **Kopka ML, Goodsell DS, Han GW, Chiu TK, Lown JW, Dickerson RE.** 1997. Defining GC-specificity in the minor groove: side-by-side binding of the di-imidazole lexitropsin to C-A-T-G-G-C-C-A-T-G. Structure **5:**1033-1046.
183. **White S, Baird EE, Dervan PB.** 1996. Effects of the A.T/T.A degeneracy of pyrrole--imidazole polyamide recognition in the minor groove of DNA. Biochemistry **35:**12532-12537.
184. **Edwards TG, Vidmar TJ, Koeller K, Bashkin JK, Fisher C.** 2013. DNA damage repair genes controlling human papillomavirus (HPV) episome levels under conditions of stability and extreme instability. PLoS One **8:**e75406.
185. **He G, Bashkin JK.** 2015. What is the antiviral potential of pyrrole-imidazole polyamides? Future Med Chem **7:**1953-1955.
186. **Koeller KJ, Harris GD, Aston K, He G, Castaneda CH, Thornton MA, Edwards TG, Wang S, Nanjunda R, Wilson WD, Fisher C, Bashkin JK.** 2014. DNA Binding Polyamides and the Importance of DNA Recognition in their use as Gene-Specific and Antiviral Agents. Med Chem (Los Angeles) **4:**338-344.
187. **Bashkin JK E, TG, Fisher C, Harris Jr GD, Koeller JK.** 2015. Antiviral Compounds and Methods for Treating Infections Caused by Double-Stranded DNA Viruses. USA.
188. **Castaneda CH, Scuderi MJ, Edwards TG, Harris GD, Jr., Dupureur CM, Koeller KJ, Fisher C, Bashkin JK.** 2016. Improved Antiviral Activity of a Polyamide Against High-Risk Human Papillomavirus Via N-Terminal Guanidinium Substitution. Medchemcomm **7:**2076-2082.
189. **Martinez TF, Phillips JW, Karanja KK, Polaczek P, Wang CM, Li BC, Campbell JL, Dervan PB.** 2014. Replication stress by Py-Im polyamides induces a non-canonical ATR-dependent checkpoint response. Nucleic Acids Res **42:**11546-11559.
190. **Qiao H, Ma C, Zhang X, Jing X, Li C, Zhao Y.** 2015. Insight into DNA Minor Groove Unspecific Binding of Pyrrole Polyamide. Bioconjug Chem **26:**2054-2061.
191. **Edelson BS, Best TP, Olenyuk B, Nickols NG, Doss RM, Foister S, Heckel A, Dervan PB.** 2004. Influence of structural variation on nuclear localization of DNA-binding polyamide-fluorophore conjugates. Nucleic Acids Res **32:**2802-2818.
192. **Kopka ML, Yoon C, Goodsell D, Pjura P, Dickerson RE.** 1985. The molecular origin of DNA-drug specificity in netropsin and distamycin. Proc Natl Acad Sci U S A **82:**1376-1380.

193. **Filippova EV, Kieser KJ, Luan CH, Wawrzak Z, Kiryukhina O, Rubin EJ, Anderson WF.** 2016. Crystal structures of the transpeptidase domain of the Mycobacterium tuberculosis penicillin-binding protein PonA1 reveal potential mechanisms of antibiotic resistance. FEBS J **283:**2206-2218.
194. **Nayak D, Panda D, Das SC, Luo M, Pattnaik AK.** 2009. Single-amino-acid alterations in a highly conserved central region of vesicular stomatitis virus N protein differentially affect the viral nucleocapsid template functions. J Virol **83:**5525-5534.
195. **Chen M, Ogino T, Banerjee AK.** 2007. Interaction of vesicular stomatitis virus P and N proteins: identification of two overlapping domains at the N terminus of P that are involved in N0-P complex formation and encapsidation of viral genome RNA. J Virol **81:**13478-13485.
196. **Filone CM, Hodges EN, Honeyman B, Bushkin GG, Boyd K, Platt A, Ni F, Strom K, Hensley L, Snyder JK, Connor JH.** 2013. Identification of a broad-spectrum inhibitor of viral RNA synthesis: validation of a prototype virus-based approach. Chem Biol **20:**424-433.
197. **Stillman EA, Rose JK, Whitt MA.** 1995. Replication and amplification of novel vesicular stomatitis virus minigenomes encoding viral structuraL-proteins. J Virol **69:**2946-2953.
198. **Horowitz S, Yesselman JD, Al-Hashimi HM, Trievel RC.** 2011. Direct evidence for methyl group coordination by carbon-oxygen hydrogen bonds in the lysine methyltransferase SET7/9. J Biol Chem **286:**18658-18663.
199. **Yesselman JD, Horowitz S, Brooks CL, 3rd, Trievel RC.** 2015. Frequent side chain methyl carbon-oxygen hydrogen bonding in proteins revealed by computational and stereochemical analysis of neutron structures. Proteins **83:**403-410.
200. **Muller G LM, Harder S.** 1996. Methyl group conformation-determining intermolecular C-H center dot center dot center dot O hydrogen bonds: Structure of N-methyl-2-pyrrolidone. Acta Crystallogr B **52:**1014-1022.
201. **Kielkopf CL, Baird EE, Dervan PB, Rees DC.** 1998. Structural basis for G.C recognition in the DNA minor groove. Nat Struct Biol **5:**104-109.
202. **Luo M.** 2012. The nucleocapsid of vesicular stomatitis virus. Sci China Life Sci **55:**291-300.
203. **Gutsche I, Desfosses A, Effantin G, Ling WL, Haupt M, Ruigrok RW, Sachse C, Schoehn G.** 2015. Structural virology. Near-atomic cryo-EM structure of the helical measles virus nucleocapsid. Science **348:**704-707.
204. **Guo Y, Wang W, Sun Y, Ma C, Wang X, Wang X, Liu P, Shen S, Li B, Lin J, Deng F, Wang H, Lou Z.** 2016. Crystal Structure of the Core Region of Hantavirus Nucleocapsid Protein Reveals the Mechanism for Ribonucleoprotein Complex Formation. J Virol **90:**1048-1061.
205. **Kirchdoerfer RN, Abelson DM, Li S, Wood MR, Saphire EO.** 2015. Assembly of the Ebola Virus Nucleoprotein from a Chaperoned VP35 Complex. Cell Rep **12:**140-149.
206. **Dong S, Yang P, Li G, Liu B, Wang W, Liu X, Xia B, Yang C, Lou Z, Guo Y, Rao Z.** 2015. Insight into the Ebola virus nucleocapsid assembly mechanism: crystal structure of Ebola virus nucleoprotein core domain at 1.8 A resolution. Protein Cell **6:**351-362.
207. **Kabsch W.** 2010. Xds. Acta Crystallogr D Biol Crystallogr **66:**125-132.
208. **Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R,**

Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr **66:**213-221.

209. **Emsley P, Cowtan K.** 2004. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr **60:**2126-2132.
210. **Schrodinger, LLC.** 2015. The PyMOL Molecular Graphics System, Version 1.8.
211. **Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA.** 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U S A **98:**10037-10041.
212. **Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE.** 2004. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem **25:**1605-1612.
213. **Baird EE, Dervan PB.** 1996. Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids. Journal of the American Chemical Society **118:**6141-6146.
214. **Lacy ER, Le NM, Price CA, Lee M, Wilson WD.** 2002. Influence of a terminal formamido group on the sequence recognition of DNA by polyamides. J Am Chem Soc **124:**2153-2163.
215. **Dupureur CM, Bashkin JK, Aston K, Koeller KJ, Gaston KR, He G.** 2012. Fluorescence assay of polyamide-DNA interactions. Anal Biochem **423:**178-183.
216. **Wang S, Nanjunda R, Aston K, Bashkin JK, Wilson WD.** 2012. Correlation of local effects of DNA sequence and position of beta-alanine inserts with polyamide-DNA complex binding affinities and kinetics. Biochemistry **51:**9796-9806.
217. **Bashkin JK, Aston K, Ramos JP, Koeller KJ, Nanjunda R, He G, Dupureur CM, David Wilson W.** 2013. Promoter scanning of the human COX-2 gene with 8-ring polyamides: unexpected weakening of polyamide-DNA binding and selectivity by replacing an internal N-Me-pyrrole with beta-alanine. Biochimie **95:**271-279.
218. **Wang S, Aston K, Koeller KJ, Harris GD, Jr., Rath NP, Bashkin JK, Wilson WD.** 2014. Modulation of DNA-polyamide interaction by beta-alanine substitutions: a study of positional effects on binding affinity, kinetics and thermodynamics. Org Biomol Chem **12:**7523-7536.
219. **Yabukarski F, Lawrence P, Tarbouriech N, Bourhis JM, Delaforge E, Jensen MR, Ruigrok RW, Blackledge M, Volchkov V, Jamin M.** 2014. Structure of Nipah virus unassembled nucleoprotein in complex with its viral chaperone. Nat Struct Mol Biol **21:**754-759.
220. **Renner M, Bertinelli M, Leyrat C, Paesen GC, Saraiva de Oliveira LF, Huiskonen JT, Grimes JM.** 2016. Nucleocapsid assembly in pneumoviruses is regulated by conformational switching of the N protein. Elife **5:**e12627.
221. **Das SC, Nayak D, Zhou Y, Pattnaik AK.** 2006. Visualization of intracellular transport of vesicular stomatitis virus nucleocapsids in living cells. J Virol **80:**6368-6377.
222. **Whelan SP, Ball LA, Barr JN, Wertz GT.** 1995. Efficient recovery of infectious vesicular stomatitis virus entirely from cDNA clones. Proc Natl Acad Sci U S A **92:**8388-8392.
223. **Harouaka D, Wertz GW.** 2012. Second-site mutations selected in transcriptional regulatory sequences compensate for engineered mutations in the vesicular stomatitis virus nucleocapsid protein. J Virol **86:**11266-11275.

224. **Rahmeh AA, Morin B, Schenk AD, Liang B, Heinrich BS, Brusic V, Walz T, Whelan SP.** 2012. Critical phosphoprotein elements that regulate polymerase architecture and function in vesicular stomatitis virus. Proc Natl Acad Sci U S A **109:**14628-14633.
225. **Raymond DD, Piper ME, Gerrard SR, Smith JL.** 2010. Structure of the Rift Valley fever virus nucleocapsid protein reveals another architecture for RNA encapsidation. Proc Natl Acad Sci U S A **107:**11769-11774.
226. **Guo Y, Wang W, Ji W, Deng M, Sun Y, Zhou H, Yang C, Deng F, Wang H, Hu Z, Lou Z, Rao Z.** 2012. Crimean-Congo hemorrhagic fever virus nucleoprotein reveals endonuclease activity in bunyaviruses. Proc Natl Acad Sci U S A **109:**5046-5051.
227. **Raymond DD, Piper ME, Gerrard SR, Skiniotis G, Smith JL.** 2012. Phleboviruses encapsidate their genomes by sequestering RNA bases. Proc Natl Acad Sci U S A **109:**19208-19213.
228. **Niu F, Shaw N, Wang YE, Jiao L, Ding W, Li X, Zhu P, Upur H, Ouyang S, Cheng G, Liu ZJ.** 2013. Structure of the Leanyer orthobunyavirus nucleoprotein-RNA complex reveals unique architecture for RNA encapsidation. Proc Natl Acad Sci U S A **110:**9054-9059.
229. **Dong H, Li P, Elliott RM, Dong C.** 2013. Structure of Schmallenberg orthobunyavirus nucleoprotein suggests a novel mechanism of genome encapsidation. J Virol **87:**5593-5601.
230. **Dong H, Li P, Bottcher B, Elliott RM, Dong C.** 2013. Crystal structure of Schmallenberg orthobunyavirus nucleoprotein-RNA complex reveals a novel RNA sequestration mechanism. RNA **19:**1129-1136.
231. **Ariza A, Tanner SJ, Walter CT, Dent KC, Shepherd DA, Wu W, Matthews SV, Hiscox JA, Green TJ, Luo M, Elliott RM, Fooks AR, Ashcroft AE, Stonehouse NJ, Ranson NA, Barr JN, Edwards TA.** 2013. Nucleocapsid protein structures from orthobunyaviruses reveal insight into ribonucleoprotein architecture and RNA polymerization. Nucleic Acids Res **41:**5912-5926.
232. **Jiao L, Ouyang S, Liang M, Niu F, Shaw N, Wu W, Ding W, Jin C, Peng Y, Zhu Y, Zhang F, Wang T, Li C, Zuo X, Luan CH, Li D, Liu ZJ.** 2013. Structure of severe fever with thrombocytopenia syndrome virus nucleocapsid protein in complex with suramin reveals therapeutic potential. J Virol **87:**6829-6839.
233. **Alayyoubi M, Leser GP, Kors CA, Lamb RA.** 2015. Structure of the paramyxovirus parainfluenza virus 5 nucleoprotein-RNA complex. Proc Natl Acad Sci U S A **112:**E1792-1799.
234. **Surtees R, Ariza A, Punch EK, Trinh CH, Dowall SD, Hewson R, Hiscox JA, Barr JN, Edwards TA.** 2015. The crystal structure of the Hazara virus nucleocapsid protein. BMC Struct Biol **15:**24.
235. **Leung DW, Borek D, Luthra P, Binning JM, Anantpadma M, Liu G, Harvey IB, Su Z, Endlich-Frazier A, Pan J, Shabman RS, Chiu W, Davey RA, Otwinowski Z, Basler CF, Amarasinghe GK.** 2015. An Intrinsically Disordered Peptide from Ebola Virus VP35 Controls Viral RNA Synthesis by Modulating Nucleoprotein-RNA Interactions. Cell Rep **11:**376-389.
236. **Reguera J, Malet H, Weber F, Cusack S.** 2013. Structural basis for encapsidation of genomic RNA by La Crosse Orthobunyavirus nucleoprotein. Proc Natl Acad Sci U S A **110:**7246-7251.

237. **Guryanov SG, Liljeroos L, Kasaragod P, Kajander T, Butcher SJ.** 2015. Crystal Structure of the Measles Virus Nucleoprotein Core in Complex with an N-Terminal Region of Phosphoprotein. J Virol **90:**2849-2857.
238. **Severin C, Terrell JR, Zengel JR, Cox R, Plemper RK, He B, Luo M.** 2016. Releasing the genomic RNA sequestered in the mumps virus nucleocapsid. J Virol doi:10.1128/JVI.01422-16.
239. **Li Z, Gabbard JD, Mooney A, Chen Z, Tompkins SM, He B.** 2013. Efficacy of parainfluenza virus 5 mutants expressing hemagglutinin from H5N1 influenza A virus in mice. J Virol **87:**9604-9609.
240. **Ryder AB, Buonocore L, Vogel L, Nachbagauer R, Krammer F, Rose JK.** 2015. A viable recombinant rhabdovirus lacking its glycoprotein gene and expressing influenza virus hemagglutinin and neuraminidase is a potent influenza vaccine. J Virol **89:**2820-2830.
241. **Tang G, Peng L, Baldwin PR, Mann DS, Jiang W, Rees I, Ludtke SJ.** 2007. EMAN2: an extensible image processing suite for electron microscopy. J Struct Biol **157:**38-46.
242. **Lawson ND, Stillman EA, Whitt MA, Rose JK.** 1995. Recombinant vesicular stomatitis viruses from DNA. Proc Natl Acad Sci U S A **92:**4477-4481.
243. **Whitt MA.** 2010. Generation of VSV pseudotypes using recombinant DeltaG-VSV for studies on virus entry, identification of entry inhibitors, and immune responses to vaccines. J Virol Methods **169:**365-374.
244. **Das SC, Pattnaik AK.** 2005. Role of the hypervariable hinge region of phosphoprotein P of vesicular stomatitis virus in viral RNA synthesis and assembly of infectious virus particles. J Virol **79:**8101-8112.
245. **Flanagan EB, Ball LA, Wertz GW.** 2000. Moving the glycoprotein gene of vesicular stomatitis virus to promoter-proximal positions accelerates and enhances the protective immune response. J Virol **74:**7895-7902.
246. **Rasmussen TB, Uttenthal A, de Stricker K, Belak S, Storgaard T.** 2003. Development of a novel quantitative real-time RT-PCR assay for the simultaneous detection of all serotypes of foot-and-mouth disease virus. Arch Virol **148:**2005-2021.
247. **Livak KJ, Schmittgen TD.** 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods **25:**402-408.
248. **Ammayappan A, Nace R, Peng KW, Russell SJ.** 2013. Neuroattenuation of vesicular stomatitis virus through picornaviral internal ribosome entry sites. J Virol **87:**3217-3228.
249. **Zhang Z, Dai W, Wang Y, Lu C, Fan H.** 2013. Analysis of synonymous codon usage patterns in torque teno sus virus 1 (TTSuV1). Arch Virol **158:**145-154.
250. **Sanjuan R, Nebot MR, Chirico N, Mansky LM, Belshaw R.** 2010. Viral mutation rates. J Virol **84:**9733-9748.
251. **Rima BK, Earle JA, Baczko K, ter Meulen V, Liebert UG, Carstens C, Carabana J, Caballero M, Celma ML, Fernandez-Munoz R.** 1997. Sequence divergence of measles virus haemagglutinin during natural evolution and adaptation to cell culture. J Gen Virol **78 (Pt 1):**97-106.
252. **Deka H, Chakraborty S.** 2016. Insights into the Usage of Nucleobase Triplets and Codon Context Pattern in Five Influenza A Virus Subtypes. J Microbiol Biotechnol **26:**1972-1982.

253. **Wang B, Yang C, Tekes G, Mueller S, Paul A, Whelan SP, Wimmer E.** 2015. Recoding of the vesicular stomatitis virus L gene by computer-aided design provides a live, attenuated vaccine candidate. MBio **6**.
254. **Friedman RA, Honig B.** 1995. A free energy analysis of nucleic acid base stacking in aqueous solution. Biophys J **69:**1528-1535.
255. **Milles S, Jensen MR, Communie G, Maurin D, Schoehn G, Ruigrok RW, Blackledge M.** 2016. Self-Assembly of Measles Virus Nucleocapsid-like Particles: Kinetics and RNA Sequence Dependence. Angew Chem Int Ed Engl **55:**9356-9360.
256. **Gagniuc P, Ionescu-Tirgoviste C.** 2012. Eukaryotic genomes may exhibit up to 10 generic classes of gene promoters. BMC Genomics **13:**512.
257. **Gagniuc P CP, Tuduce R, Ionescu-Tirgoviste C, Gavrila L.** 2012. DNA Patterns and Evolutionary Signatures Obtained through Kappa Index of Coincidence. Rev Roum Sci Techn - Electrotechn et Energ **57:**100-109.
258. **Gumpper RH, Li W, Castaneda CH, Scuderi MJ, Bashkin JK, Luo M.** 2018. A Polyamide Inhibits Replication of Vesicular Stomatitis Virus by Targeting RNA in the Nucleocapsid. J Virol **92**.
259. **Duffy S.** 2018. Why are RNA virus mutation rates so damn high? PLoS Biol **16:**e3000003.
260. **Fitzsimmons WJ, Woods RJ, McCrone JT, Woodman A, Arnold JJ, Yennawar M, Evans R, Cameron CE, Lauring AS.** 2018. A speed-fidelity trade-off determines the mutation rate and virulence of an RNA virus. PLoS Biol **16:**e2006459.
261. **Butt AM, Nasrullah I, Tong Y.** 2014. Genome-wide analysis of codon usage and influencing factors in chikungunya viruses. PLoS One **9:**e90905.
262. **Lauring AS, Acevedo A, Cooper SB, Andino R.** 2012. Codon usage determines the mutational robustness, evolutionary capacity, and virulence of an RNA virus. Cell Host Microbe **12:**623-632.
263. **Ma XX, Feng YP, Liu JL, Zhao YQ, Chen L, Guo PH, Guo JZ, Ma ZR.** 2014. The characteristics of synonymous codon usage in the initial and terminal translation regions of encephalomyocarditis virus. Acta Virol **58:**86-91.
264. **Ma MR, Hui L, Wang ML, Tang Y, Chang YW, Jia QH, Wang XH, Yan W, Ha XQ.** 2014. Overall codon usage pattern of enterovirus 71. Genet Mol Res **13:**336-343.
265. **Tulloch F, Atkinson NJ, Evans DJ, Ryan MD, Simmonds P.** 2014. RNA virus attenuation by codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide frequencies. Elife **3:**e04531.
266. **Cheng X, Virk N, Chen W, Ji S, Ji S, Sun Y, Wu X.** 2013. CpG usage in RNA viruses: data and hypotheses. PLoS One **8:**e74109.
267. **Kumar CS, Kumar S.** 2017. Synonymous codon usage of genes in polymerase complex of Newcastle disease virus. J Basic Microbiol **57:**481-503.
268. **Kumar N, Bera BC, Greenbaum BD, Bhatia S, Sood R, Selvaraj P, Anand T, Tripathi BN, Virmani N.** 2016. Revelation of Influencing Factors in Overall Codon Usage Bias of Equine Influenza Viruses. PLoS One **11:**e0154376.
269. **Cristina J, Moreno P, Moratorio G, Musto H.** 2015. Genome-wide analysis of codon usage bias in Ebolavirus. Virus Res **196:**87-93.
270. **Ball LA, Pringle CR, Flanagan B, Perepelitsa VP, Wertz GW.** 1999. Phenotypic consequences of rearranging the P, M, and G genes of vesicular stomatitis virus. J Virol **73:**4705-4712.

271. **Gun L, Haixian P, Yumiao R, Han T, Jingqi L, Liguang Z.** 2018. Codon usage characteristics of PB2 gene in influenza A H7N9 virus from different host species. Infect Genet Evol **65:**430-435.
272. **Zhang X, Cai Y, Zhai X, Liu J, Zhao W, Ji S, Su S, Zhou J.** 2018. Comprehensive Analysis of Codon Usage on Rabies Virus and Other Lyssaviruses. Int J Mol Sci **19**.
273. **Hunter J.** 2007. Matplotlib: A 2D graphics environment. Computing in Science & Engineering **9:**90-95.

APPENDICES

Appendix A

Appendix A.1

$2^{-\Delta\Delta C_T}$ Values for Minigenome

Comparison:	$2^{-\Delta\Delta C_T}$
UMSL1011:DMSO(control)	0.33
DMSO(control):Mock(control)	0.88
UMSL1011:Mock(control)	0.29

Appendix A.2

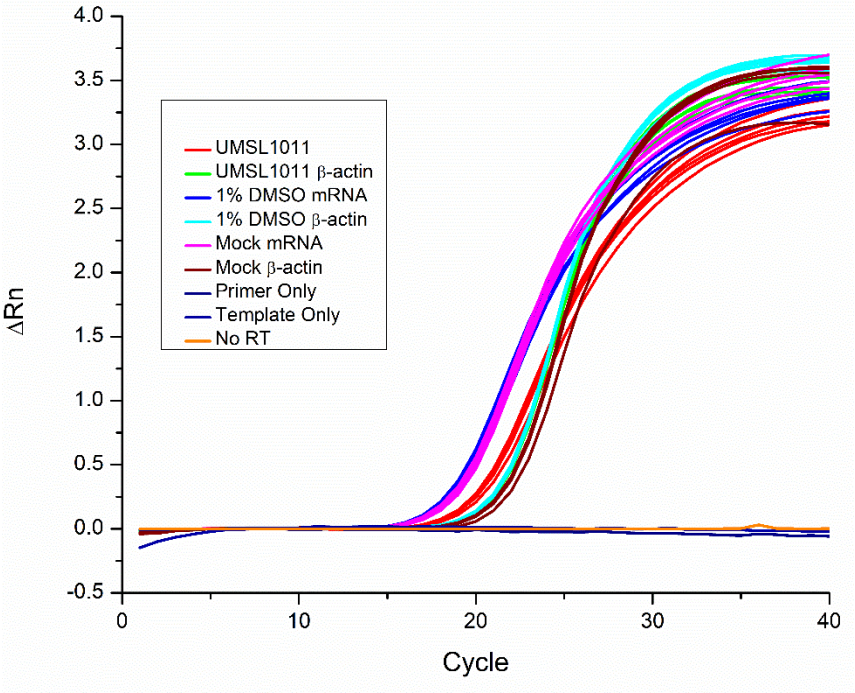


Figure A.2 Amplification plots of the qPCR in the minigenome assays in **Figure 3.6**. This illustrates that the controls do not have any amplification even when compared to the experimental conditions.

Appendix A.3

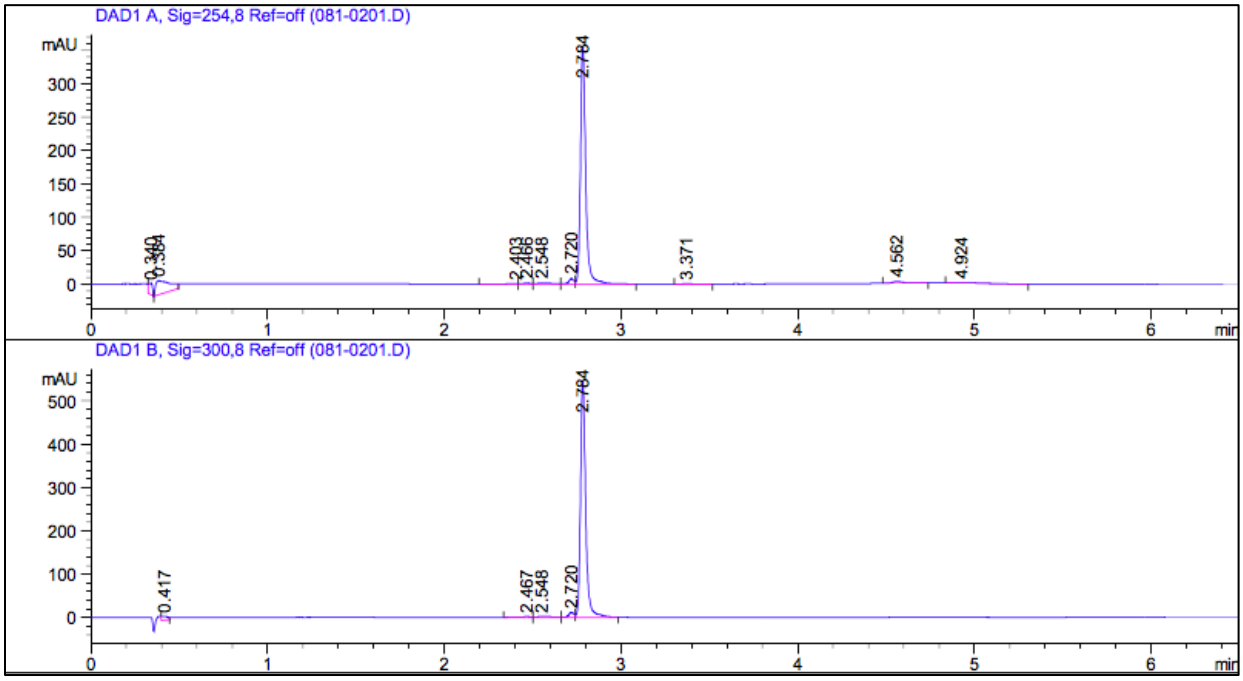
Polyamide Characterization

UMSL1013 Im-Py-Py-Py-γ-Py-Py-Im-Py-β-Dp (3 TFA):

^{1}H NMR (600 MHz, DMSO-d_6) being obtained now.

HRMS (ESI) calculated for $C_{58}H_{71}N_{21}O_{10}$ $[MH]^+$, 1221.5684, found, 1221.5702.

Reverse Phase HPLC purity (trace shown): 94%.

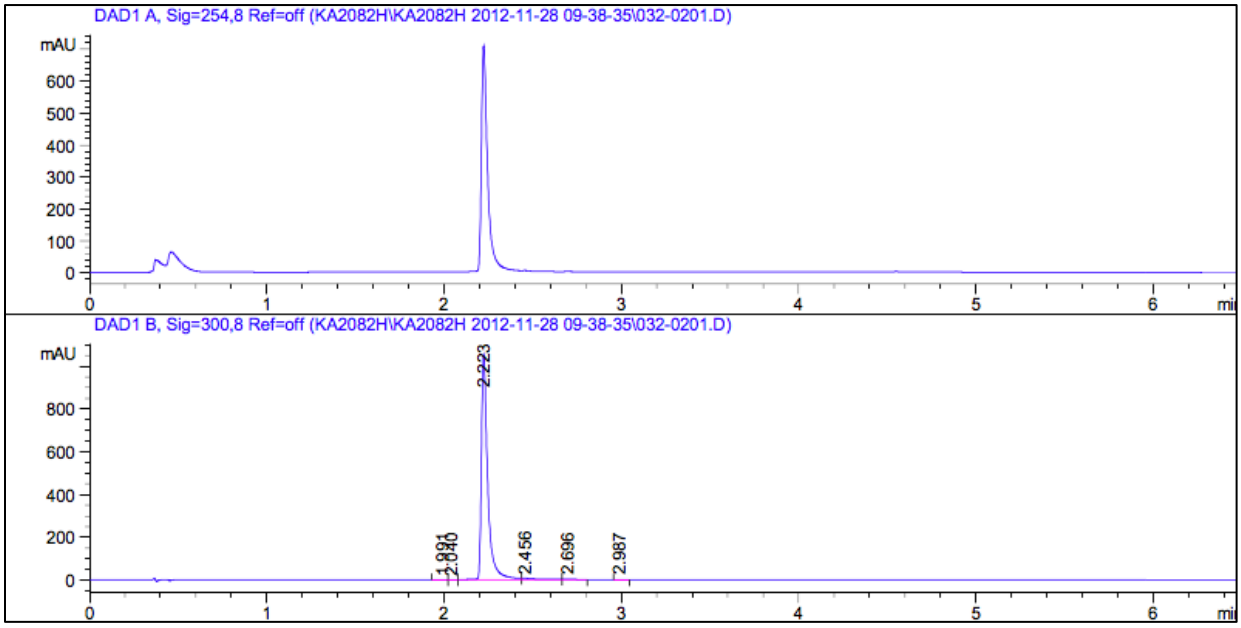


UMSL2082 Im-Im-Py-γ-Py-Py-Py-β-Ta (4 TFA):

^{1}H NMR (600 MHz, DMSO-d_6) δ = 7.54 (s, 1 H), 7.42 (s, 1 H), 7.20 -7.18 (dd, 1 H), 7.13 (s, 1 H), 7.07 (d, 1 H), 7.01 (s, 1 H), 6.96 (d, 1 H), 6.85-6.84 (m, 2 H), 3.98-3.97 (m, 7 H), 3.80-3.77 (m, 14 H), 3.48 (19 H, H_2O), 3.37-3.35 (m, 3 H), 3.20-2.99 (m, 12 H), 2.84-2.81 (m, 3 H), 2.73-2.71 (s, 4 H), 2.35-2.32 (t, 2 H), 2.27-2.25 (t, 2 H), 1.90-1.86 (m, 3 H), 1.79-1.73 (m, 5 H).

HRMS (ESI) calculated for $C_{48}H_{64}N_{18}O_8$ $[M]^+$, 1020.5154, found, 1020.5101.

Reverse Phase HPLC purity (trace shown): 98%.



HP