Structural Studies of RNA Packaging by Phlebovirus Nucleocapsid Proteins and RNA Capping by Flavivirus Methyltransferases

by

Donald Damian Raymond

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biological Chemistry)
in The University of Michigan
2012

Doctoral Committee

Professor Janet L Smith, Chair

Professor Alexander J Ninfa

Professor John J Tesmer

Associate Professor Raymond C Trievel

Assistant Professor Georgios Skiniotis

TO THE RAYMOND FAMILY

ACKNOWLEDGEMENTS

First, I would like to acknowledge my mentor and friend Dr. Janet Smith for her guidance and support over the years. During my time in Janet's lab, she taught me how to think critically, write professionally and present my research clearly. For that I am deeply appreciative. Many thanks to my thesis committee, Drs. Alex Ninfa, Raymond Trievel, John Tesmer and Yiorgo Skiniotis. Their helpful suggestions and advice kept me focused, as I tended to drift from one project to the next. I would like to thank Drs. Mary Piper and Sonja Gerrard for a thrilling collaboration, and insightful discussions about all things viral. I also thank the Richard Kuhn lab at Purdue University for their assistance with materials for the flavivirus project.

I would like to acknowledge and thank my outstanding lab mates — past and present — for their support, advice and expertise. We live and worked together for many years, and I will always cherish our time together.

I would like to thank the faculty and staff of the Department of Biological Chemistry for their support and encouragements over the years and for creating a stimulating environment where young scientists can flourish.

Finally, I thank the members of the extended Raymond family for their encouragement and support over their years.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

BDV	Borna disease virus
DENV	Dengue fever virus
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
FLUVA	Influenza A virus
FP	Fluorescence polarization
GMP	Guanosine monophosphate
G_N & G_C	Glycoproteins
GTase	Guanylyltransferase
GTP	Guanosine triphosphate
HRSV	Human respiratory syncytial virus
IPTG	Isopropyl β-D-1-thiogalactopyranoside
JEV	Japanese encephalitis virus
kV	Kilovolts
L	Large segment
LGTV	Langat virus
LVMT	Langat virus methyltransferase
M	Medium segment
Mg^{2+}	magnesium
mP	Millipolarization
MTase	Methyltransferase
N	Nucleocapsid protein
N_{core}	Nucleocapsid protein core
NCS	Non-crystallographic symmetry
NS3	Non-structural protein 3

NS5	Non-structural protein 5
NSV	Negative-sense RNA virus
PCR	Polymerase chain reaction
RABV	Rabies virus
RdRp	RNA-dependent RNA polymerase
recN	Recombinant N
RMSD	Root mean squared deviation
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
RVFV	Rift Valley fever virus
S	Small segment
S200	Supexdex 200
SAH	*S*-Adenosyl homocysteine
SAM	*S*-Adenosyl methionine
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SeMet	Selenomethionyl
SUMO	Small Ubiquitin-like Modifier
TBEV	Tick-borne encephalitis virus
TCEP	Tris(2-carboxyethyl)phosphine
TLS	Translation-libration-screw
TOSV	Toscana virus
VSV	Vesicular stomatitis virus
w/v	Weight-in-volume
WNV	West Niles virus
YFV	Yellow fever virus

ABSTRACT

Arboviruses are maintained in nature through a complex life cycle involving blood-sucking insects and susceptible vertebrate hosts. Arboviruses have evolved mechanisms to conceal their RNA genomes from the cellular antiviral response. In this study, I investigated genome packaging in phleboviruses and RNA capping in flaviviruses.

Crystal structures of phlebovirus N-RNA complexes reveal a surprising diversity of binding geometries where variation in the RNA length and the number of nucleotides per N subunit gives rise to tetrameric, pentameric and hexameric structures. RVFV N binds RNA by sequestering all nucleotide bases in an RNA-binding slot with the sugar-phosphate backbone facing the solvent. The structures reveal multiple conformations of an N-terminal helical arm. The flexibility of the arm allows for the asymmetry of the N-RNA multimers and the asymmetric architecture of the RNP. The crystal structures reveal a common building block consisting of the core domain of an N subunit, four RNA nucleotides, and the helical arm of an adjacent subunit (RNA-N_{core}-arm). Together with direct binding measurements showing that N binds nucleic acids with equal affinity regardless of length or sequence, these results provide a model for sequence-independent base sequestration by N, explain the observed structure of phlebovirus RNP, and elucidate how phlebovirus N protects the RNA genome from the cellular antiviral response.

High resolution structures of Langat virus methyltransferase (LVMT) in apo form and in complex with GTP and GpppA cap analogues provide a detailed view of GTP and RNA cap binding to the MTase GTP-binding site and suggest possible interactions that confer guanine specificity. An RNA-cap bound structure provides the first view of an RNA cap fully extended in the active site cleft and shows the correct orientation of the RNA cap prior to ribose 2′-O methylation at the first

nucleotide. Biochemical characterization of the LVMT guanylyltransferase activity using radiolabeled GTP showed the formation of the MTase-GMP adduct, which is required for the capping reaction. These results enhance our overall understanding of the cap formation process in flaviviruses.

CHAPTER 1
INTRODUCTION

Arthropod-borne viruses

Arthropod-borne viruses (arboviruses) are maintained in nature through a complex life cycle involving haematophagous insects and vertebrate hosts. Arboviruses primarily belong to RNA virus families (*Flaviviridae*, *Togaviridae*, *Bunyaviridae*, *Rhabdoviridae* and *Reoviridae*), with the exception of the African swine fever virus, which has a double-stranded DNA genome. Arboviruses primarily infect non-human vertebrates, but some are associated with human diseases (Fig. 1.1). An RNA genome is beneficial to arboviruses due to the high error rate of the RNA-dependent RNA polymerase (RdRp) and the high levels of viral replication (1, 2). The high rate of spontaneous mutations enables arboviruses to infect new host organisms and spread to new environments (2).

Mosquitos transmit the majority of arboviruses, including dengue fever virus (DENV, *Flaviviridae*), West Niles virus (WNV, *Flaviviridae*), yellow fever virus (YFV, *Flaviviridae*), Rift Valley fever virus (RVFV, *Bunyaviridae*), Japanese encephalitis virus (JEV, *Flaviviridae*), California encephalitis virus (CEV, *Bunyaviridae*) and many others (3). Other human arboviral pathogens, including tick-borne encephalitis virus (TBEV, *Flaviviridae*), Colorado tick fever virus (*Reoviridae*), Louping Ill virus (*Flaviviridae)* and Langat viruses (LGTV, *Flaviviridae*), are transmitted by infected ticks (4). Sandflies, biting midges and many other blood-sucking insects are also capable of spreading arboviruses (5).

Most arbovirus infections are enzootic and primarily occur in wild or ruminant animals. Humans become infected tangentially and do not function as primary hosts in the transmission cycles of the viruses (Fig. 1.2). Enzootic episodes usually occur during the summer or rainy season when environmental conditions are favorable for the propagation of the arthropod vector. During epidemics, arboviral infections often decimate wild bird and mammal populations (3, 5, 6).

Arboviruses are a significant threat to global human public health. More than 100 arboviruses are capable of causing human disease, and new viruses are continuously being discovered (7, 8). Arboviruses continue to expand their range

geographically and more than 14,000 species of blood-sucking insects are capable of transmitting arboviruses (5, 9). Even arboviruses that do not cause significant human disease have significant economic impact by causing disease in livestock populations (8).

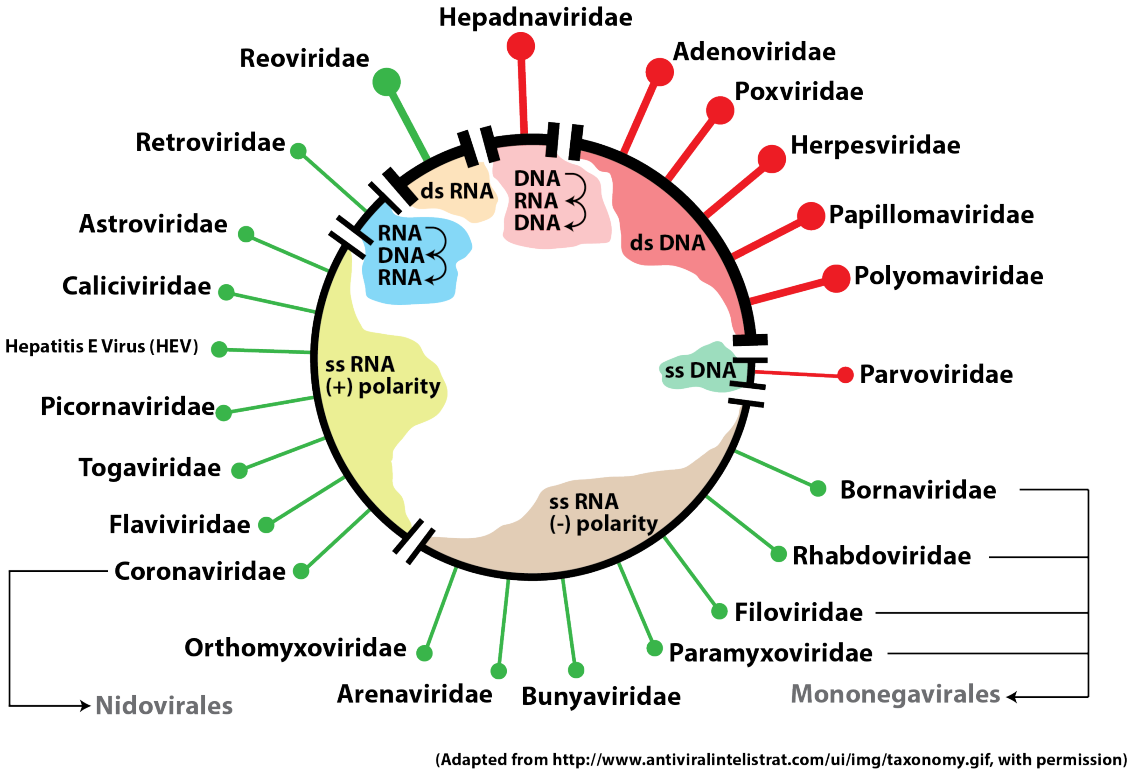


Figure 1.1 ***Virus families associated with human disease.***
The spikes represent virus families, which are classified by genome type (ss for single-stranded and ds for double-stranded genomes). Curved arrows represent conversion of viral genome from DNA to RNA or vice versa during an intermediate step in the virus replicative cycle.

With the exception of YFV, no vaccines or effective therapeutics exist to prevent or treat arbovirus infections (10, 11). Developing and disseminating vaccines and therapeutics for these viruses has been exceedingly challenging (11-13). The plasticity of arbovirus RNA genomes enables the viruses to rapidly evolve into new immunologically distinct virus strains, which require a new vaccines (2).

Currently, the most effective method of managing arbovirus epidemics is through vector control (13). In the early 20th century, pesticides like DDT were used heavily to control arboviruses. Though successful in eliminating DENV and YFV from the United States by the 1950s, DDT use was banned in 1972 due to the environmental

impacts of its indiscriminate use, particularly on birds (14). Presently, there is a resurgence of DENV infections in the southern United States, which is compounded by the emergence of pesticide-resistant vectors throughout the country (13). Therefore, further work is required to discover weaknesses in the viral replicative cycles to exploit for developing therapeutics.

To identify potential targets for developing therapeutics, I investigated the mechanisms that two arboviruses use to conceal their RNA genomes from the host antiviral response. First, I discuss RNA packaging in the Rift Valley fever virus (RVFV). The RVFV nucleocapsid protein (N) packages the RNA genome by producing N-RNA complexes, which protect the genome from degradation. Next, I discuss the Langat virus RNA-capping enzyme. The flavivirus RNA-capping enzyme caps and methylates the 5′ ends of newly synthesized genomic RNA to prevent RNA turnover by host RNases, mimic cellular mRNA for translation by the ribosomes and to evade the host immune response.

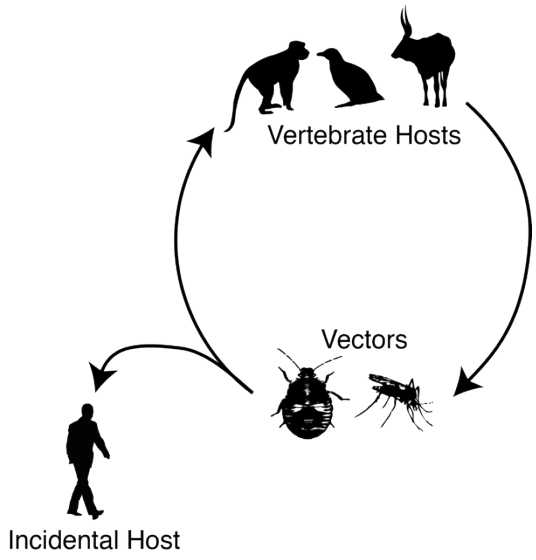


Figure 1.2 ***Transmission cycle of a typical arbovirus*****.**
Arboviruses are maintained in nature through enzootic cycles between haematophagous insects and vertebrate hosts like rodents, birds and other non-human primates. Human infections arise from spill over from zoonotic cycles when vector populations are high or humans enter enzootic habitats.

Rift Valley fever virus

Rift Valley fever virus (RVFV), a *Phlebovirus* in the *Bunyaviridae* family, is an emerging arthropod-borne pathogen first isolated in Kenya in the 1930s, but currently found throughout sub-Saharan Africa, Madagascar, northern Africa, and recently the Arabian peninsula (15-20). The virus infects livestock and humans and is primarily transmitted by infected mosquitoes, with epidemics and epizootics appearing most often after periods of heavy rainfall (17, 21). The virus can also be transmitted by handling infected meat products or by aerosol. The ability to transmit RVFV by aerosol allows for the possibility of weaponizing the virus (22).

RVFV causes Rift Valley fever disease, which leads to a high death rate in ruminant animals, especially among cattle and sheep, with the young animals being particularly susceptible (15, 23-25). The virus causes spontaneous abortion in nearly all infected pregnant animals (19, 21, 23). Symptoms of RVFV infections in humans include encephalitis, hemorrhagic fever and retinal hemorrhage, which can lead to permanent vision loss (26). The disease has a fatality rate of up to 10% of confirmed cases depending on the viral strain (27). The long summer months in the United States are perfect for mosquito breeding, and the mosquito vectors capable of amplifying and spreading RVFV are present throughout the country.

Genomic organization

Like all bunyaviruses, the RVFV replicative cycle occurs in the cytoplasm of infected cells (Fig 1.3)(28). RVFV contains three negative-sense, single-stranded RNA segments designated as large (L), medium (M) and small (S) (Fig. 1.4)(28, 29). The L segment contains ~6400 nucleotides and codes for a 240 kDa RNA-dependent RNA polymerase (RdRp), which is responsible for replicating and transcribing the genome. The M segment contains ~3800 nucleotides and codes for a 132 kDa polypeptide, which is post-translationally cleaved to yield G_N and G_C glycoproteins. The glycoproteins have ectodomains, which are responsible for host recognition and fusion. The S segment contains ~1200 nucleotides and codes for a 27 kDa nucleocapsid protein (N), which packages and protects the RNA genome.

The interaction between the RNA genome and N is essential for virus viability. Before new virus particles can assemble, each RNA genome segment is completely encapsidated by N to produce a protein-RNA complex called the ribonucleoprotein (RNP) (30). When the genome is packaged into RNP, it is protected from ribonuclease degradation and from the cellular antiviral response, which warned neighboring cells of viral presence by secreting interferon. The RVFV RdRp requires RNP for both replication and transcription and will not replicate naked RNA (31). Furthermore, the gylcoproteins (G_N and G_C) interact with N to recruit RNP to the site of viral assembly (31, 32).

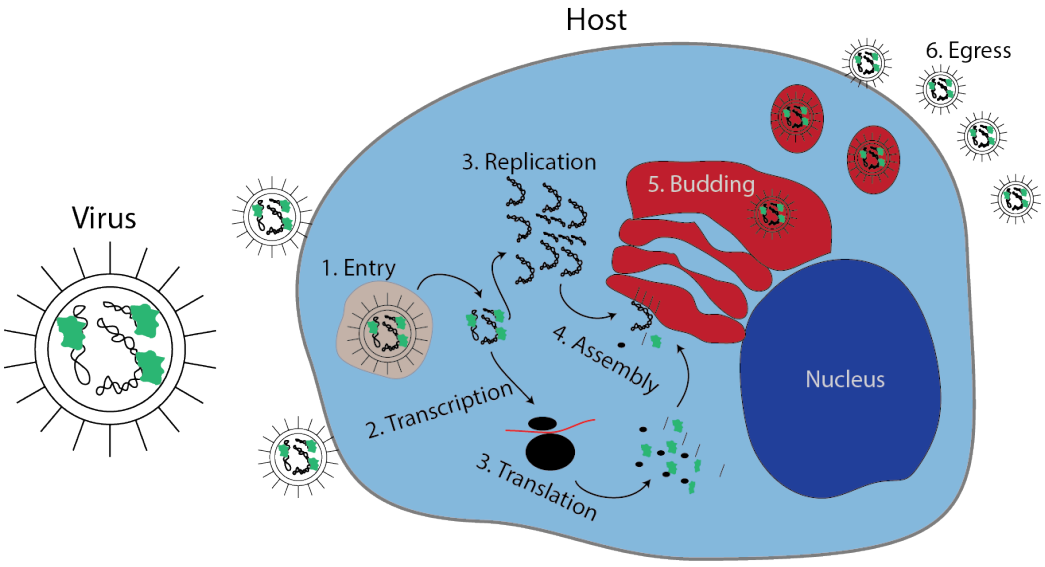


Figure 1.3 ***The phlebovirus replicative cycle.***
(1) The virus enters the host cell by receptor-mediated endocytosis. (2) mRNA transcription by RdRp. (3) Genome replication by RdRp and translation by ribosomes. (4) Virus assembly at the Golgi apparatus membrane. (5) Budding from the Golgi apparatus. (6) Release of newly packaged viruses.

Molecular details about phlebovirus RNA genome packaging or the architecture of the N-RNA complex are unknown. Phlebovirus N are significantly smaller and have no sequence homology to N from extensively studied negative-sense RNA viruses such as influenza, rabies, measles and Ebola viruses, which have helical RNP (33). The goal of this study was to determine how differences in phlebovirus N translated to differences in the phlebovirus genome packaging strategy.

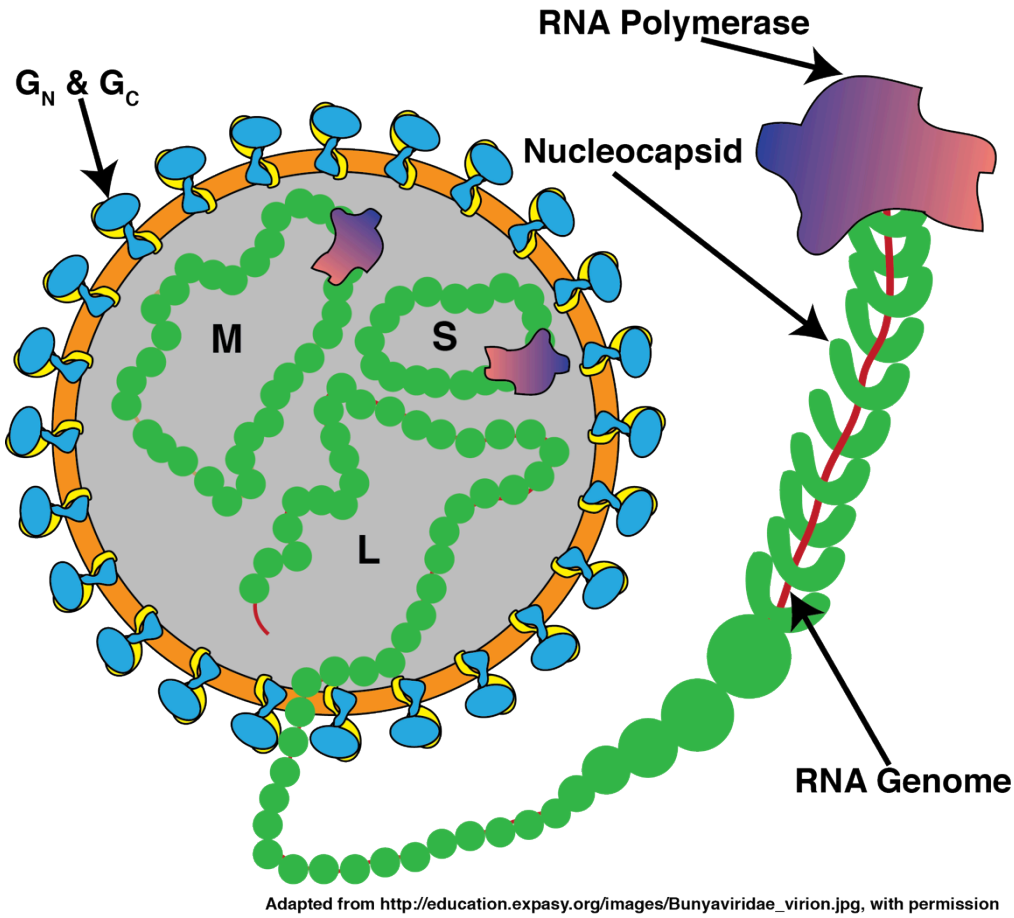


Figure 1.4 ***Schematic of a RVFV particle.***
L is the large genomic segment, M is the medium genomic segment and S is the small genomic segment. G_N and G_C are the glycoproteins responsible for host recognition and entry. RdRp molecules are purple, N is green, the lipid membrane is orange, and the RNA genome is red. The average diameter of RVFV particles is 95 nm.

Flaviviruses

The flavivirus genus comprises more than 70 different viral species distributed throughout the world (34). About half of flaviviruses are responsible for some of the most important arthropod-borne human diseases. Representative members of the flavivirus genus include dengue virus (DENV), yellow fever virus (YFV), West Nile virus (WNV), Japanese encephalitis virus (JEV), and tick borne encephalitis virus (TBEV) (35). YFV, the prototype of the flavivirus genus, infects about 200,000 people a year, while 2.5-3 billion of the world's population is at risk of DENV infection (36-38).

Genome organization and transcription

Flaviviruses are characterized by an ~11 kB single-stranded positive-sense RNA genome with a cap at the 5′ end and no 3′ polyadenylation (39). Two non-coding regions flank a single open reading frame encoding an ~3400 residue polypeptide (Fig. 1.5), which is co-translationally inserted into the endoplasmic reticulum membrane during translation (40, 41). The polypeptide is processed by a viral protease in the cytoplasm and a host protease in the ER lumen to produce at least ten mature proteins; three structural and seven non-structural proteins (archC-prM-E-NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5) (Fig. 1.5)(42-44). The structural proteins are located at the N-terminus of the polypeptide and are used for assembling new virus particles (39). The non-structural proteins (NS1-5) are located at the C-terminus of the polypeptide and are responsible for viral genome replication and immune evasion (45-47).

Flavivirus Replicase Complex

NS3 is a multifunctional protein with an N-terminal serine protease domain and a C-terminal domain with NTPase, RTPase and 3′-5′ RNA-helicase activities (48-51). The protease is active as a complex with its cofactor NS2B (43, 49). The C-terminal helicase domain was first identified by sequence analysis as a member of the DEAH/D box superfamily II helicase family due to the presence of seven conserved sequence motifs including Walker A and Walker B (52-54). It has since been shown

in vitro to unwind RNA duplexes with a 3′ overhang (55, 56). The NTPase and RTPase activities for several flaviviruses have been shown in other *in vitro* experiments (48, 51, 57-59). The role of the RNA helicase domain is to unwind duplex RNA before or after replication and to prepare the 5′ end of the genome for capping. Crystal structures of helicases from DENV, YFV, and Murry Valley viruses revealed that the helicase is structurally conserved within flaviviruses (60-62).

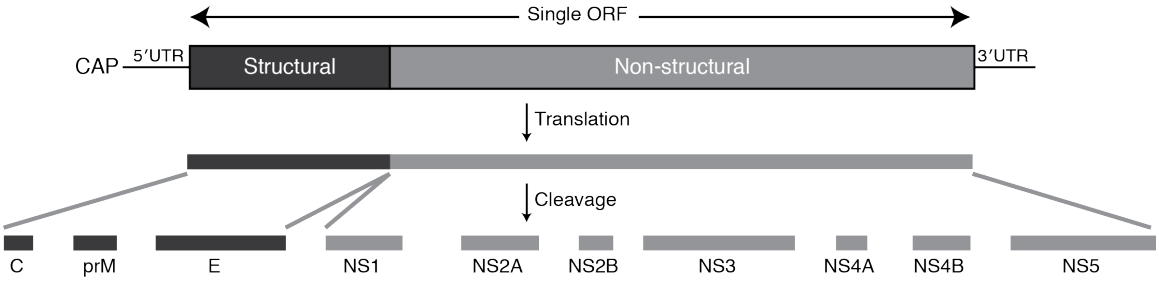


Figure 1.5 ***Flavivirus genome and polypeptide.***
A typical flavivirus genome consists of a single-stranded positive sense RNA with a 5′ cap and no polyA tail. The single ORF produces a ~3400 amino acid polypeptide, which is proteolytically processed by viral and host proteases.

NS5 is another multifunctional viral protein with an N-terminal *S*-adenosyl-L-methionine (SAM) dependent methyltransferase (MTase) and a C-terminal RNA-dependent RNA-polymerase (RdRp) (63, 64). The enzymes were first annotated after sequencing the YFV genome (39). A SAM-binding motif indicated that the N-terminal domain is a SAM-dependent methyltransferase. Seven conserved sequence motifs found in all polymerases and four specific to RdRp domains indicated that the C-terminal domain is an RdRp (63). The flaviviruses MTase domains have been shown to perform both the N7 and 2′-O methylation reactions required for the formation of a mature 5′ cap (65-67). Several groups have demonstrated that recombinant flavivirus RdRp can catalyze the synthesis of a complimentary strand from a RNA template (68-74).

NS3 and NS5 form a viral replicase complex responsible for replicating, capping and methylating the genome (69). NS3 and NS5 from DENV and YFV have been shown to directly interact *in vivo* using infected cells, and *in vitro* using a recombinant vaccinia virus system (75-77). *In vivo* immunofluorescence experiments performed using WNV corroborated the interaction between the NS3 and NS5 (78).

Biochemical studies showed that recombinant NS5 enhances the NTPase activity of NS3 (79). Furthermore, NS3 catalyzes the RTPase activity required to prepare the 5′ end of the genome for capping, whereas NS5 catalyzes the guanyltransfer and both the N7 and 2′-O methylation required to complete the 5′ cap (51, 66, 67).

Because an RNA cap is essential for virus viability, probing RNA cap formation and maturation by flavivirus proteins is important for anti-viral drug design. The flavivirus capping machinery differs from that of eukaryotes and would be an ideal drug target (11, 80). In this study, we investigated possible structural differences between the mosquito- and tick-borne flavivirus capping enymes. We solved the structure of Langat virus MTase (LVMT) in complex with SAH, GTP and cap analogues and performed structural comparisons with DENV, YFV and WNV MTase domains. We also investigated the guanylyltransferase activity of LVMT.

Outline of thesis

In Chapter 2, I describe the crystal structure of monomeric RVFV RNA-free N in the inactive conformation and the electron microscopy analysis of the irregular, non-helical structure of authentic and reconstituted phlebovirus RNP. Our analyses reveal that phlebovirus encapsidated genome is of substantially different organization than in other negative-sense RNA virus families.

In Chapter 3, I describe the crystal structure of reconstituted N-RNA oligomers. The N-RNA crystal structures reveal a new paradigm for non-specific RNA binding. They showed that phleboviruses use a base-sequestration mechanism where bases are non-specifically inserted into an RNA-binding slot. This is the first example of a nucleic-acid binding protein that binds the bases (as opposed to the phosphate backbone) in a non-specific manner.

In Chapter 4, I describe our structural analysis of flavivirus RNA capping and cap methylation by a tick-borne flaviviruses. High-resolution crystal structures of Langat virus methyltransferase in complex with capped RNA analogues provide insights into cap formation and methylation in flaviviruses. We also demonstrate the guanylyltransferase activity of the flavivirus MTase domain.

In Chapter 5, I summarize our findings and propose future directions for both projects.

CHAPTER 2

STRUCTURE OF RVFV RNA-FREE N

Summary

Rift Valley fever virus (RVFV) is a negative-sense RNA virus (genus Phlebovirus, family Bunyaviridae) that infects livestock and humans and is endemic to sub-Saharan Africa. Like all negative-sense RNA viruses, the segmented RNA genome of RVFV is encapsidated by a nucleocapsid protein (N). The 1.93-Å crystal structure of RVFV N and electron micrographs of ribonucleoprotein (RNP) reveal an encapsidated genome of substantially different organization than in other negative-sense RNA virus families. The RNP polymer, viewed in electron micrographs of both virus RNP and RNP reconstituted from purified N with a defined RNA, has an extended structure without helical symmetry. N-RNA species of ~100-kDa apparent molecular weight and heterogeneous composition were obtained by exhaustive ribonuclease treatment of virus RNP, by recombinant expression of N, and by reconstitution from purified N and an RNA oligomer. RNA-free N, obtained by denaturation and refolding, has a novel all-helical fold that is compact and well ordered at both the N and C termini. Unlike N of other negative-sense RNA viruses, monomeric RVFV N has no positively charged surface cleft for RNA binding and no protruding termini or loops to stabilize a defined N-RNA oligomer or RNP helix. A potential protein interaction site was identified in a conserved hydrophobic pocket. The nonhelical appearance of phlebovirus RNP, the heterogeneous ~100-kDa N-RNA multimer, and the fold of N differ substantially from the RNP and N of other negative-sense RNA virus families and provide valuable insights into the structure of the encapsidated phlebovirus genome.

Publication note: Research described in Chapter 2 was originally published in *Proceedings of the National Academy of Sciences.* 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.

Introduction

RNA viruses cause a myriad of human and animal diseases, including measles, poliomyelitis, rabies, the common cold, dengue fever, and Rift Valley fever. Despite tremendous diversity among RNA viruses, all must package a protected RNA genome into virus particles. RNA viruses protect their genome in one of two ways, providing either a protein shell or a protein coat for the genome (81). The process is generally known by the term encapsidation, however functionally and structurally encapsidation takes a variety of forms. Most positive-sense RNA and double-stranded RNA viruses place their genome within a protein shell, known as a capsid. By contrast, the negative-sense RNA viruses encapsidate their genome by coating the length of the RNA with a nucleocapsid protein (N). Although capsid proteins and N both bind RNA, the resulting RNA-protein complexes differ, and it is not possible to make generalizations about the proteins involved in encapsidation across all RNA virus families.

Rift Valley fever is a mosquito- and aerosol-borne disease of livestock and humans in sub-Saharan Africa, and is caused by Rift Valley fever virus (RVFV). Rift Valley fever in domestic ruminants results in abortion and high rates of mortality, especially among young animals (82). In humans, Rift Valley fever is typically a self-limited febrile illness, although severe disease, such as hemorrhagic fever and encephalitis, occurs in a small percentage of cases (82). No effective therapeutics exist for treating Rift Valley fever. RVFV has a membrane envelope and a genome comprised of three negative-sense RNA segments, termed small (S), medium (M) and large (L) (82). It belongs to the *Phlebovirus* genus in the *Bunyaviridae* family. As with all negative-sense RNA viruses, the genome is completely encapsidated by a nucleocapsid protein (N). The N of RVFV is a 27-kDa protein encoded by the S segment. Encapsidation of RVFV genomic RNA, as with all negative-sense RNA viruses, plays an essential role in multiple steps within the replicative cycle including transcription and replication by the RNA-directed RNA polymerase (RdRp) and packaging of genome into virions (83). RVFV N is thought to interact

with the viral RdRp because N is essential to replication and transcription (84). N also plays a role in virus assembly through interactions with the viral envelope glycoproteins (G_N and G_C) (32).

Bunyavirus N binds single-stranded RNA (ssRNA) non-specifically (85, 86), although some N may have a preference for specific viral RNA sequences (87-89). Studies on animal viruses within the *Bunyaviridae* family found that encapsidated RNA is resistant to disruption by high salt and ribonuclease treatment (87, 89, 90). Despite the common property of tight, non-specific binding to single-stranded RNA, homology of N within the *Bunyaviridae* family is not apparent from sequence data, as N from different genera appear unrelated. However, within a genus, the N are clearly homologous. When RVFV N is compared across the *Phlebovirus* genus, the amino acid identity ranges from 50% to 59% , and is 36% for Uukuniemi virus , the most divergent clade within the *Phlebovirus* genus. The high degree of sequence identity indicates that the phlebovirus N have similar structures and form similar RNPs. Additionally, the phlebovirus N are distantly related to the N of the *Tenuivirus*, a genus of negative-sense RNA viral plant pathogens with worldwide distribution (91). Otherwise, the phlebovirus N appear unrelated to N of other negative-sense RNA viruses.

Structures are available for N from several negative-sense RNA viruses, including influenza A virus (FLUVA (92)), rabies virus (RABV (93)), human respiratory syncytial virus (HRSV (94)), vesicular stomatitis virus (VSV (95)) and Borna disease virus (BDV (96)). For some of these, ribonucleoprotein (RNP) complexes have been visualized by crystallography or electron microscopy (FLUA (97), RABV (93), HRSV (94), VSV (95). The crystallized RNP oligomers are high-order ring structures formed by specific contacts of loops or chain termini of neighboring N subunits. In some viruses, the number of subunits in the ring matches the helical repeat of the RNP polymers (94). None of the structurally characterized N is from the *Bunyaviridae* family and none has detectable homology with the phlebovirus N. Early EM studies of encapsidated bunyavirus genomes reveal a non-condensed, macro-circular form that appears to lack symmetry (98, 99). Nevertheless, negative-

sense RNA viruses are assumed to have condensed helical structures based on micrographs of RNP from several virus families (94, 100-103).

Here we report the 1.93 Å crystal structure of recombinant RVFV N and views of two forms of RVFV RNP by electron microscopy. N has a novel protein fold that differs substantially from N of other negative-sense RNA viruses. The re-folded, recombinant N forms stable multimeric N-RNA complexes of similar appearance to N-RNA multimers released from virus RNP by exhaustive ribonuclease treatment. The N-RNA multimer is heterogeneous with 4-7 N subunits and has an apparent molecular weight of 100 kDa. Authentic virus RNP and RNP reconstituted from re-folded N and defined RNA have a similar non-helical appearance and similar ribonuclease resistance.

Experimental Procedures

Plasmid Construction

All plasmids were generated by Mary Piper using standard molecular cloning techniques and were confirmed by sequencing. pTrRVFV-SΔNSs::GFP, pRdRp-Amp and pN-Amp has been described previously (104, 105). To generate the protein expression construct for N, the gene was amplified using Phusion polymerase (New England Biolabs), pTrRVFV-SΔNSs::GFP as template, and primers 5′-GACGTGGGTCTCGAGGTATGGACAACTATC-AAGAGCTTG-3′ and 5′-CTCGAGTTAG-GCTGCTGTCTTGTAAGCCTG-3′. The PCR product was cloned into pCRII-Blunt-TOPO (Invitrogen). Digestion with BsaI and XhoI liberated the N ORF, which was subsequently ligated into pSUMO (Life sensors, Inc.), thus producing pIPER1. For the RdRp transcription assay, the plasmids pN and pRdRp were constructed by subcloning the open reading frames from pN-Amp and pRdRp-Amp into pVAX1 (Invitrogen) using the HindIII/EcoRI and BamHI/NotI sites, respectively. pSTrRVFV-SΔNSs::hRLuc was derived from pTrRVFV-SΔNSs::GFP in several steps. First, the GFP gene was released by digestion with EcoRV, followed by ligation with a Renilla luciferase gene that was flanked with EcoRV sites. The resulting plasmid was then subcloned into pSMART HC Kan (Lucigen). pSTrRVFV-SΔNΔNSs::hRLuc was derived from pSTrRVFV-SΔNSs::hRLuc by removing the 237 nucleotide SmaI fragment from the N gene. The RVFV N mutants, W125A, R178Q, and R178E, were generated using Phusion polymerase, pN as template, and primer pairs 5′-GTCTTG-AGTGAGGCGCTTCCTGTCACTG-3′ and 5′-CAGTGACAGGAAGCGCCTCACTCAAGAC-3′, 5′-CTGCAGTTCTCCCAGGTCATCAACCCA-3′ and 5′-TGGGTTGATGACCTGGGAGAAC-TGCAG-3′, and 5′-CTGCAGTTCTCCGAGGTCATCAACCCA-3′ and 5′-TGGGTTGATGAC-CTCGGAGAACTGCAG-3′, respectively.

Cells and virus

BSR-T7/5 cells expressing the T7 RNA polymerase were a generous gift of Dr. K. Conzelmann (Max-von Pettenkofer-Institut, Munchen, Germany). BSR-T7/5 cells were subsequently cloned by limiting-dilution and the resulting clonal lines were

screened using the RVFV transcription assay. The C3 clone of the BSR T7/5 line was used for all experiments. The cells were grown in Dulbecco's Modified Eagle Medium (Invitrogen) supplemented with 10% fetal calf serum, and 1 mg/mL geneticin. RVFV ZH548 MP12 vaccine strain was a generous gift of Dr. R. Tesh (World Reference Center of Emerging Viruses and Arboviruses). (Mary Piper and Sonja Gerrard performed these experiments).

Protein Expression

pIPER1 was transformed into *E. coli* strain BL21 AI (Invitrogen) containing the pRARE2 plasmid (Novagen) (106) and grown in 1 L of TB media (12 g tryptone, 24 g yeast extract, 2.31 g KH_2PO_4(monobasic), 12.5 g K_2HPO_4 (dibasic), 40 mL glycerol) containing 35 µg/mL chloramphenicol and 50 µg/mL kanamycin at 37˚C until OD_{600}=1.0. The temperature was reduced to 20˚C, and expression was induced after 1 hr by addition of 4 mL 50% w/v arabinose and isopropyl-β-D-thiogalactopyranose (IPTG) to a final concentration of 0.4 mM. The cultures were incubated 12 hr at 20˚C and cells were harvested by centrifugation. All purification steps were carried out at 4˚C unless otherwise noted.

Selenomethionine labeling for phasing

To produce the selenomethioninyl (SeMet) variant of N, 50-mL cultures were grown in rich media and cells were harvested and added to SelenoMet Medium Base supplemented with SelenoMet Nutrient Mix (Athena Enzyme Systems) and 100 mg/L of D, L-SeMet to give an initial OD_{600} of 0.3. Cultures were grown to OD_{600}=0.5, incubated 1 hr at 20˚C, and induced with 0.4 mM IPTG. SeMet N was purified identically to the wild type.

Protein purification under denaturing conditions

Initial attempts to purify the recombinant protein under native conditions resulted in protein bound to heterologous *E. coli* RNA (Fig. 2.1a), so the protein was purified under denaturing conditions and refolded (107). Briefly, cell pellets were resuspended in 35 mL lysis buffer (50 mM Na-phosphate pH 7.8, 1 M NaCl, 20 mM

imidazole, 5% glycerol), lysed by sonication, and centrifuged at 27,000 x g for 45 minutes. The supernatant was loaded onto a 5-mL HiTrap chelating column (GE Healthcare) pre-equilibrated with lysis buffer. After a wash with 10 column volumes of lysis buffer, the protein was unfolded with a linear gradient of 0-8 M urea in lysis buffer over 5 column volumes. After a wash with 5 column volumes of lysis buffer with 8 M urea, the protein was refolded with a linear gradient of 8-0 M urea in lysis buffer over 10 column volumes followed by a wash with 10 column volumes of lysis buffer. The protein was eluted with a linear gradient of 20-500 mM imidazole in lysis buffer. Fractions containing SUMO-N, as determined by 12% SDS-PAGE, were pooled and dialyzed 1 hr against 1 L dialysis buffer (50 mM Na-phosphate pH 7.8, 1 M NaCl, 5% glycerol).

SUMO-fusion removal

The His_6-SUMO fusion was cleaved by incubation of SUMO-N at 4˚C overnight with His-tagged SUMO-protease at a final concentration of 1:1000 (protease:protein) and dialysis was continued with fresh buffer for 16 hr. The expression plasmid for SUMO protease was a kind gift of C. Lima, Memorial Sloan-Kettering Cancer Center (108). The proteolysis mixture was loaded on a 5-mL HiTrap column pre-equilibrated with lysis buffer, and cleaved protein was washed from the column with lysis buffer. Finally, the SUMO-protease and His_6-SUMO fusion were eluted from the column with 3 column volumes of 500 mM imidazole in lysis buffer.

Size exclusion chromatography

N was concentrated using Centriprep-10 (Millipore), and subjected to size exclusion chromatography by a HiLoad 16/60 Superdex 75 gel filtration column (Amersham) pre-equilibrated with storage buffer (20 mM Tris pH 7.8, 0.5 M NaCl, 10% glycerol). The fractions corresponding to the N monomer peak (Fig. 2.1b) were pooled and concentrated to 13 mg/mL using Centriprep-10. Purified protein was flash-frozen in liquid N_2 and stored at -80˚C. Typical 1-L cultures yielded 20 mg of purified, RNA-free N (Fig 2.1c, d).

Protein purification under native conditions

Recombinant N-RNA multimer was purified identically to refolded RNA-free N except RNase A (Invitrogen) was added at each stage of purification, there was no refolding step with urea, and a HiLoad 16/60 Superdex 200 gel filtration column was used for the final purification step (Fig. 2.1a, b). Fractions from the 100-kDa peak corresponding to the N-RNA multimer were pooled, concentrated, flash frozen, and stored at -80˚C.

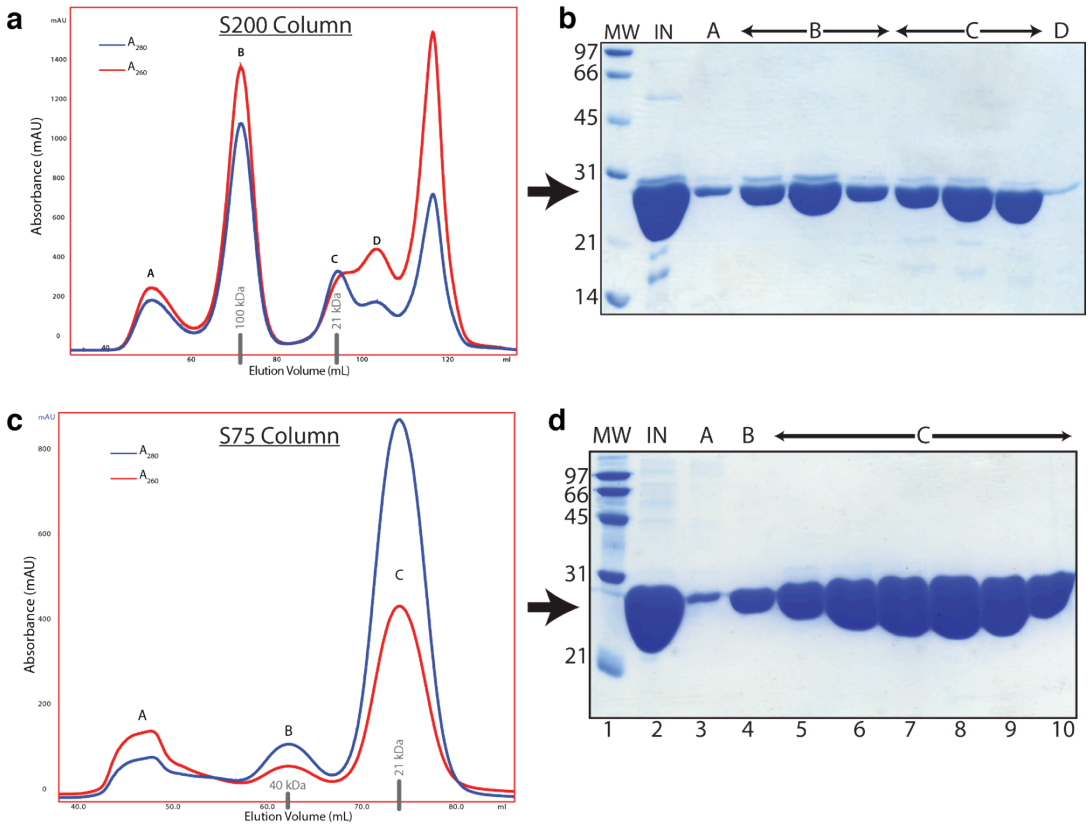


Figure 2.1 ***Purification of recombinant RVFV N.***
(**a**) Preparative S200 gel-filtration chromatogram of recombinant N after purification by Ni-affinity chromatography and cleavage of the SUMO fusion partner. Peak A is the void volume, peak B is the N-RNA multimer, peak C is the N monomer, and peaks D and E are predominantly RNA. The red and blue traces represent absorption at 260 nm and 280 nm, respectively. (**b**) SDS-PAGE of fractions from the chromatogram shown in (**a**). Lane 1: molecular weight markers in kDa, lane 2: input sample, lanes 3–9: fractions from peaks A-C, as labeled. (**c**) Preparative S75 gel-filtration chromatogram of refolded RVFV N. Peak A is the void volume, peak B is the N dimer, and peak C is the N monomer. Even after denaturation, some N is still bound to RNA and elutes as a multimer. (**d**) SDS-PAGE of fractions from the chromatogram shown in (**c**). Lane 1: molecular weight markers in kDa, lane 2: input sample, lanes 3–10: fractions from peaks A–C, as labeled.

RNA-free N Crystallization

Prior to crystallization, N was dialyzed against crystallization buffer (20 mM Tris pH 7.8, 250 mM NaCl). N was crystallized at 4˚C by hanging drop vapor diffusion from a 1:1 mixture of protein (10 mg/mL N in crystallization buffer) and well solution (26% PEG 3350, Na/K phosphate pH 5.5). Optimal crystals were obtained after 2 weeks. The crystals were cryo-protected by soaking in well solution with the addition of 10% glycerol, harvested into loops, and frozen by plunging into liquid N_2.

Data Collection and Structure Determination

Diffraction data were collected at 100 K on GM/CA beamline 23ID-D (native) and 23ID-B (SeMet) at the Advanced Photon Source (APS), Argonne National Laboratory (Argonne, IL).

Although the crystals appeared single, all diffracted in two lattices, which complicated data processing. The SeMet data were processed using HKL2000 (109), with which images could be indexed on the stronger lattice by using only data beyond 4-Å spacings for initial indexing (Tables 2.1-2). Data from crystals of wild type N were processed with iMOSFLM (110, 111), which was able to index with no exclusions, and were scaled with SCALA (Table 2.3) (112). SOLVE (113) and RESOLVE (114) were used for initial phasing using a two-wavelength MAD data set from one crystal of SeMet protein. 36 of the 48 Se sites were located and used for MAD phasing (phasing-power = 0.9, initial FOM = 0.35), followed by density modification phase refinement with four-fold averaging and automated modeling (60% of main chain) in PHENIX (115). Modeling was completed manually using Coot (116). Refinement was performed using REFMAC5 with non-crystallographic (NCS) restrains (117, 118). Individual isotropic thermal parameters were refined during each round of refinement. The structure was solved from triclinic crystals with four N polypeptides in the asymmetric units. The final model is complete except for residues 17 in chain A, 18 in chain B, 16-19 in chain C, and 16-19 and 28-30 in chain D. All residues are in favored regions of the Ramachandran plot except for Ile209 in all four chains, whose backbone conformation is well supported by

density (Fig. 2.2). The structure was validated using MolProbity (119), PyMOL (120) was used for generating figures and for molecular superpositions, sequence alignments were performed using ClustalW (121), the APBS plugin (122) in PyMOL was used to calculate electrostatic surface potential, and ESPript (123) was used for secondary structure assignment. Structure based homology models were generated using the MMM server (124) and optimized using YASARA (http://www.yasara.org). Conservation scores were calculated by the ConSurf server (125).

Table 2.1 ***HK2000 scaling summary for RNA-free N at Inflection***

Shell limit	Lower Angstrom	Upper	Average I	Average error	Norm. stat.	Linear Chi**2	Square R-fac	R-fac
50.00	5.25		259.5	22.2	8.0	2.511	0.139	0.208
5.25	4.17		248.4	21.5	7.9	1.713	0.125	0.188
4.17	3.64		223.3	19.8	7.7	1.493	0.133	0.218
3.64	3.31		148.6	13.8	6.1	1.331	0.141	0.211
3.31	3.07		90.7	9.3	4.7	1.358	0.177	0.310
3.07	2.89		54.8	6.4	3.8	1.317	0.222	0.423
2.89	2.75		42.4	5.7	3.5	1.278	0.255	0.533
2.75	2.63		31.1	4.9	3.2	1.271	0.303	0.721
2.63	2.53		23.6	4.5	3.1	1.257	0.350	0.000
2.53	2.44		20.2	4.8	3.5	1.198	0.385	0.860
All reflections			115.6	11.4	5.2	1.477	0.160	0.226

Table 2.2 ***HK2000 scaling summary for RNA-free N at peak***

Shell limit	Lower Angstrom	Upper	Average I	Average error	Norm. stat.	Linear Chi**2	Square R-fac	R-fac
50.00	5.25		364.0	25.7	10.3	1.530	0.114	0.268
5.25	4.17		328.7	24.0	10.5	1.284	0.114	0.287
4.17	3.64		287.6	21.9	10.2	1.275	0.127	0.344
3.64	3.31		185.5	15.4	8.0	1.231	0.140	0.315
3.31	3.07		111.2	10.8	6.3	1.139	0.185	0.448
3.07	2.89		65.6	8.2	5.2	1.301	0.242	0.637
2.89	2.75		49.4	7.5	4.8	1.130	0.281	0.702
2.75	2.63		35.9	7.0	4.5	1.284	0.352	0.000
2.63	2.53		26.8	6.9	4.4	1.351	0.410	0.872
2.53	2.44		22.0	7.7	5.4	1.546	0.458	0.842
All reflections			148.8	13.6	7.0	1.303	0.154	0.309

Table 2.3 ***SCALA scaling summary for native RNA-free N***

	Overall	InnerShell	OuterShell
Low resolution limit	48.32	48.32	2.03
High resolution limit	1.93	6.09	1.93
Rmerge	0.087	0.062	0.432
Rmerge in top intensity bin	0.066	-	-
Rmeas (within I+/I-)	0.101	0.072	0.501
Rmeas (all I+ & I-)	0.101	0.072	0.501
Rpim (within I+/I-)	0.051	0.037	0.253
Rpim (all I+ & I-)	0.051	0.037	0.253
Fractional partial bias	-0.047	-0.053	-0.131
Total number of observations	347126	10790	49300
Total number unique	87736	2767	12642
Mean((I)/sd(I))	9.6	18.0	3.1
Completeness	97.1	96.9	95.7
Multiplicity	4.0	3.9	3.9

Table 2.4 ***Crystallization summary*****.**

	SeMet		Native
Diffraction data			
Space group	*P*1		*P*1
Unit cell lengths (Å) a, b, c	67.0, 69.5 80.5		67.1 69.6 80.6
Unit cell angles (°) α, β, γ	82.6, 70.2, 61.4		78.4 69.7 60.9
X-ray source	APS 23 ID-D		APS 23 ID-B
Wavelength (Å)	0.9794	0.9796	1.0333
d_{min} (Å)	2.44(2.53-2.44)[a]	2.44(2.53-2.44)	1.93(2.03-1.93)
Unique reflections	44,975	44,817	87,736
R_{merge}[b] %	0.16 (0.38)	0.15 (0.45)	0.08 (0.43)
Avg I/σ_I	10.1 (4.2)	14.2 (2.8)	9.6 (3.1)
Completeness (%)	96.6 (81.7)	97.0 (86.6)	97.1 (95.7)
Average redundancy	3.9 (3.8)	3.9 (3.8)	4.0 (3.9)
Refinement			
Data range (Å)	50.0 - 2.44		46.5 - 1.93
Reflections	41,208		83,342
R/R_{free}[c]	0.221/0.279		0.211/0.254
RMSD bond lengths (Å)	0.008		0.013
RMSD bond angles (º)	1.091		1.293
Avg. protein B-factor (Å^2)	32.5		29.9
Avg. solvent B-factor (Å^2)	37.4		35
Ramachandran agreement[d]			
Allowed (%)	98		99.6
Disallowed (%)	2		0.4
Protein atoms	6969		7565
Water molecules	191		616
Glycerol	0		4
PDB Code			3LYF

[a] Values in parentheses are for the outermost shell of data.
[b] $R_{merge} = \sum|I_i - \langle I \rangle| / \sum I_i$, where I_i is the intensity of the i^{th} observation and $\langle I \rangle$ is the mean intensity. Sums are taken over all reflections. R_{merge} values include anomalous scattering.
[c] $R = \sum||F_o| - |F_c|| / \sum|F_o|$. R_{free} is calculated for a random 5% subset of the data that was excluded from refinement.
[d] Calculated with MolProbity (119).

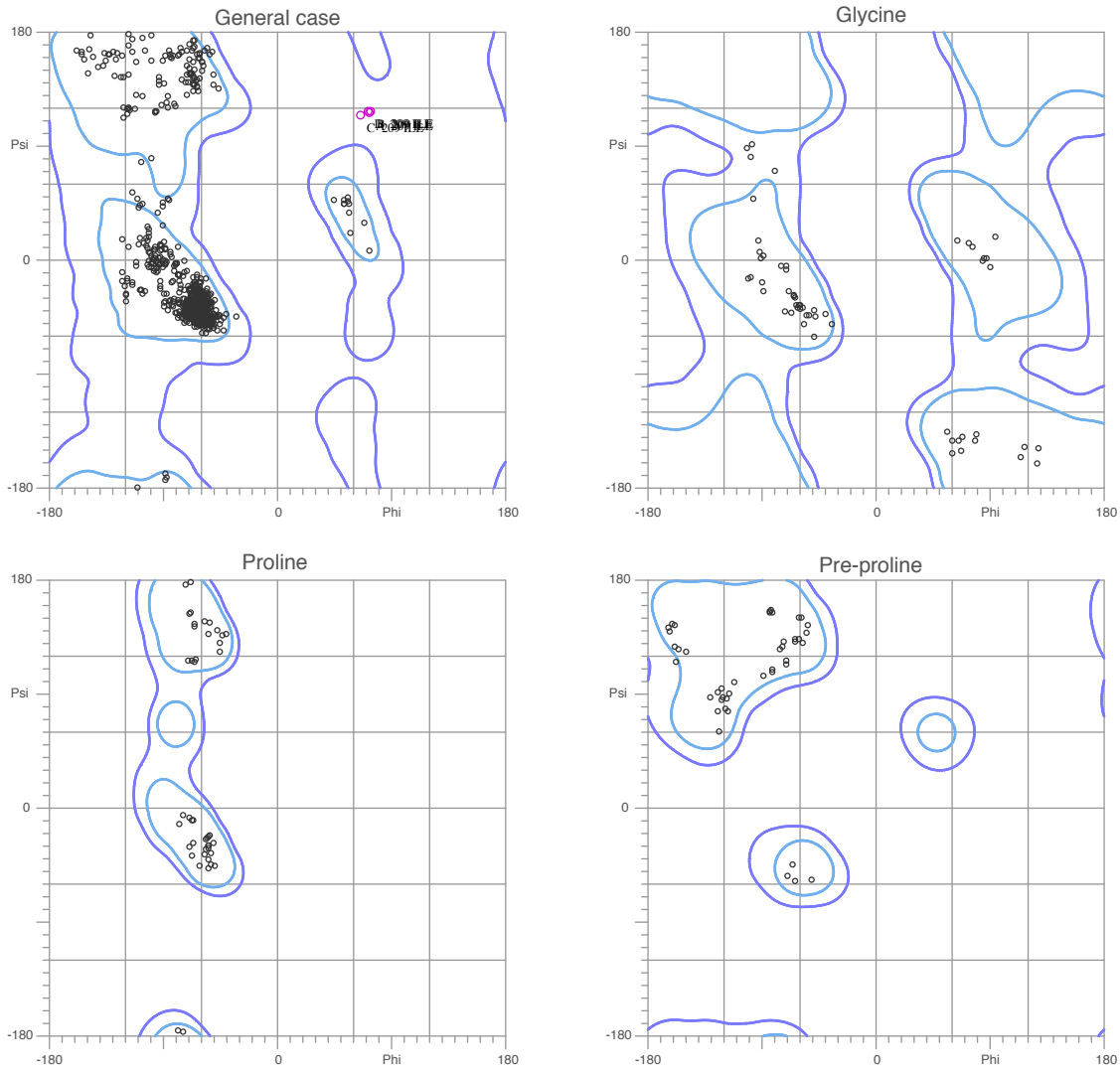


Figure 2.2 ***Ramachandran plot for RNA-free N.***
As output from MolProbity, 99.6% of residues were in the allowed regions.

Authentic RVFV RNP preparation

Encapsidated genomes or ribonucleoparticles (RNPs) were purified by Sonja Gerrard from BSR-T7/5 cells infected with the RVFV MP-12 strain. Cells and supernatants were harvested when cells started to show cytopathic effect and frozen at -80˚C. The cell slurry was thawed on ice, octyl-β-D-glucopyranoside was added to a final concentration of 1%, and the mixture was incubated on ice for 1 hr. Insoluble material was removed by centrifugation at 1,100g for 15 min at 4˚C. The supernatant was then centrifuged at 53,000g in an SW28 rotor for 16 h at 4˚C. The pellet was resuspended in TNE (10 mM Tris, pH 7.5, 100 mM NaCl, 1 mM EDTA)

with 5% sucrose. Insoluble material was removed by centrifugation at 16,000g for 5 min. The soluble material was then layered on top of a discontinuous 30-40-50% CsCl gradient that was then centrifuged at 53,000 x g in an SW28 rotor for 16 h at 4℃. A visible opaque band was observed at the 40-50% CsCl interface. Fractions were taken from the bottom of the gradient and analysed for the presence of N by ELISA. Fractions containing N corresponded to the visible band observed at the 40-50% CsCl interface. These fractions were pooled and dialyzed against 1X TNE then analysed by SDS-PAGE followed by colloidal Coomassie staining or immunoblot with polyclonal antibodies to N and RdRp (126) (Fig. 2.3).

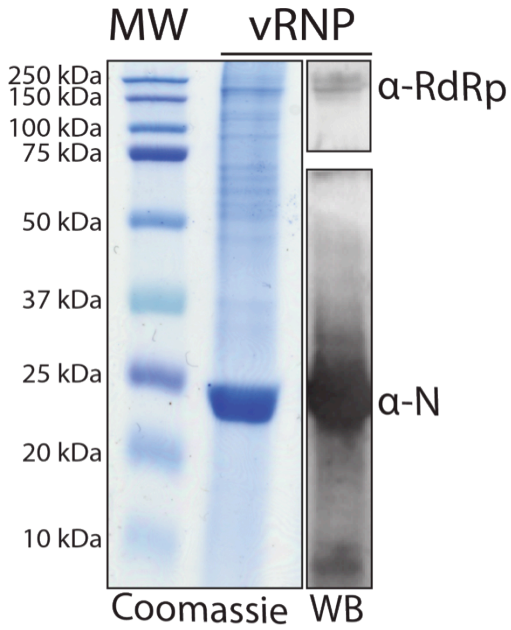


Figure 2.3 ***SDS-PAGE of virus RNP*****.**
Duplicate gels were stained with Coomassie (left panel) and blotted with polyclonal antibodies to the RVFV RdRp (top right panel) and the RVFV N (bottom right panel). N dominates the Coomassie-stained gel, but the RdRp is also visible.

RNP and N–RNA reconstitution

Reconstituted RNP was generated by incubating RNA-free N with a single-stranded 684-nt RNA (127) at a ratio of 300:1 (N:RNA) for 30 min at room temperature. N-RNA_x multimers were reconstituted by incubating RNA-free N with 10 or 25-base poly(U)-RNA at a ratio of 6:1 (N:RNA) for 30 min at room temperature.

Prior to imaging, samples of freshly prepared virus RNP and reconstituted RNP were treated with RNase A (Invitrogen) for 16 hr at room temperature. To generate an N-RNA multimer from virus RNP, the sample was sheared with a 26 G½ needle (Becton Dickinson & Co.) and treated with RNase A at 35 ˚C for 72 hours.

Electron microscopy and image processing

For electron microscopy, samples were prepared using conventional negative staining protocols (128). Briefly, 3.5 µL of sample was pipetted onto a glow-discharged carbon-coated grid and stained with 0.75% (w/v) uranyl formate. Imaging was performed at room temperature with a Morgagni 268(D) transmission electron microscope (FEI Company) equipped with a tungsten filament operated at an acceleration voltage of 100 kV and a mounted Orius SC200W CCD camera (Gatan Inc.). Magnifications were 24,628X for multimer samples and 30,461X for RNPs. For classification and averaging, 1019 particles of multimer released from virus RNP, 782 particles of recombinant N-RNA multimer, 1352 particles of reconstituted N-RNA_{25} multimer and 411 N-RNA_{10} multimer were interactively selected from micrographs using EMAN Boxer (129). Reference-free alignment and classifications into 10-25 classes for each sample were performed with SPIDER (130)(25).

Results

Protein oligomeric state in solution

Purification of recombinant N under native conditions, including exhaustive ribonuclease treatment, resulted in a discrete complex of the protein and *E. coli* nucleic acid as determined by the ratio of absorbances at 260 nm and 280 nm (Fig. 2.1a & b). The protein could not be separated from nucleic acid under native conditions using high salt concentrations and pH extremes. The recombinant complex had an apparent mass of 100 kDa by size-exclusion chromatography. Nucleic acid was extracted from the recombinant RVFV N by denaturation and then treated with either deoxyribonuclease or ribonuclease. The nucleic acid was sensitive only to ribonuclease treatment, demonstrating that N was bound to RNA (Fig. 2.4). The formation of a nonspecific ribonucleoprotein (RNP) complex between recombinant RNA-binding proteins and *E. coli* RNA is not uncommon (85, 94). Crystal structures of RABV (93), VSV (88) and HRSV (94) RNPs were solved from RNPs bound to *E. coli* RNA, however no crystals were obtained using the recombinant RVFV RNPs. We therefore used denaturation to obtain RNA-free recN for crystallization. After purification from RNA and re-folding, N was predominantly a monomer of apparent molecular mass 21 kDa, with about 10% as a dimer (Fig. 2.1b).

Cross-Linking of Authentic Virus RNPs and Recombinant N with RNA

We tested whether the RNA-free N could interact with RNA similarly to N in viral RNPs. Purified viral RNPs, re-folded recN, and reconstituted N-RNA multimer (recN-RNA) were cross-linked with a homo-bifunctional amine-reactive cross-linker and then separated by electrophoresis under denaturing conditions (Fig. 2.5). In the absence of cross-linker (Fig. 2.5, lanes 1, 4 & 9), N from all samples migrated as a monomer. When viral RNPs were exposed to increasing amounts of cross-linker, the monomer band decreased in intensity and four higher molecular weight complexes appeared (Fig. 2.5, lanes 2 & 3). Cross-linking of RNA-free, re-folded recN resulted in predominant monomer and minor dimer species (Fig. 2.5, lanes 5-8), consistent

with the behavior of recN in size-exclusion chromatography (Fig. 2.1b). In contrast, when recN-RNA was cross-linked, many species of higher molecular weight were observed (Fig. 2.5, lanes 10-13). The number of N within the dominant cross-linked species created from recN-RNA and virus RNPs (Fig. 2.5, lane 11 and Fig 1A, lane 3) was estimated to be 2, 4, 6, and 10 based on an apparent molecular weight of 25 kDa for N. Thus, re-folded recN behaves similarly to viral N in its ability to bind RNA and to form multimeric complexes. The cross-linked species formed by both viral RNPs and recN-RNA appear primarily as multiples of two, consistent with a previously reported dimeric association of N (131).

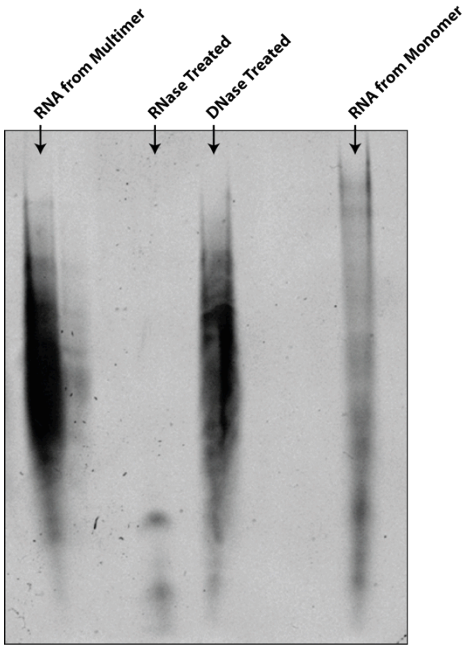


Figure 2.4 ***Nuclease sensitivity of nucleic acid extracted from RVFV N multimers.***
Nucleic acid was extracted from purified recombinant N multimer with an equal-volume mixture of phenol and chloroform. The extracted RNA was treated with RNase or DNase, separated on a denaturing urea-polyacrylamide gel and visualized using SYBR Green. (Courtesy of Mary Piper)

Electron Microscopy of Authentic Virus RNPs and Reconstituted RNPs

We compared RNPs isolated from virus with RNPs reconstituted from re-folded N and a large single-stranded RNA. Freshly prepared samples (Fig. 2.3) were viewed by negative-stain electron microscopy before and after overnight ribonuclease digestion at room temperature. Both authentic virus RNP and reconstituted RNP were ribonuclease-resistant and had a remarkably similar string-like appearance

(Figs 2.6a-b, & Fig. 2.7). No helical symmetry was apparent in either sample. The appearance of phlebovirus RNP is strikingly different from images of similarly prepared RNP from other negative-sense RNA viruses, which have obvious helical symmetry (94, 100-103).

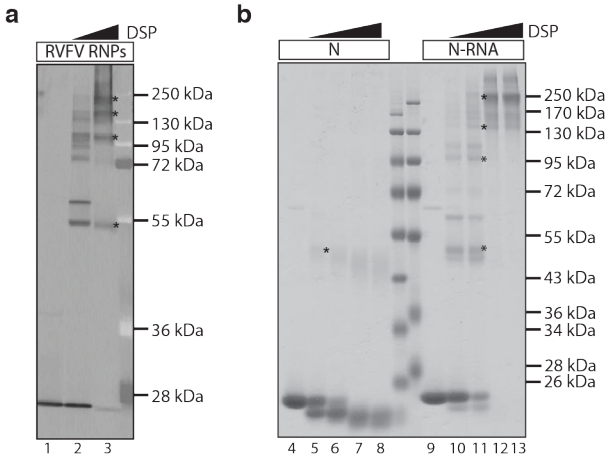


Figure 2.5 ***Similar multimer complexes of viral RNPs and purified RNA-free N bound to RNA.*** (**a**) Viral RNP. Purified RVFV RNPs were cross-linked with 0.0, 5.0, or 20.0 mM DSP and analyzed by immuno-blot. Asterisks indicate predominant cross-linked species. Molecular weight markers are in the rightmost lane. (**b**) Recombinant N. N or N bound to U25 single-stranded RNA (N-RNA) was cross-linked using 0.0, 1.0, 5.0, 10.0, or 20.0 mM DSP, separated by SDS-PAGE, and visualized with colloidal Coomassie stain. The dominant cross-linked species are indicated by asterisks. (Courtesy of Mary Piper)

Heterogeneous Multimeric N-RNA Complexes

Aggressive ribonuclease digestion (3 days at 37°C) released multimers from the virus RNP (Figs 2.6a & 2.7b). As determined by EM, this particle is similar in appearance to the ribonuclease-resistant, recombinant N-RNA multimer purified from *E. coli* (Figs 2.7c & 2.8b) and to multimers reconstituted from re-folded N and defined RNA oligomers (Figs 2.6c-d & 2.7d). Remarkably, the multimers are of similar size distribution (10-12 nm diameter) regardless of the source of RNA. The N-RNA multimers from all sources appear heterogeneous, with 4-7 bright objects per particle (Fig. 2.7b-d). The heterogeneity of the recombinant N-RNA multimer explains its inability to crystallize. The recombinant and reconstituted multimers have an apparent molecular weight of 90-100 kDa by size-exclusion chromatography (Fig. 2.9). Thus, we conclude that the N-RNA multimers contain 4-7 N subunits and an unknown amount of RNA.

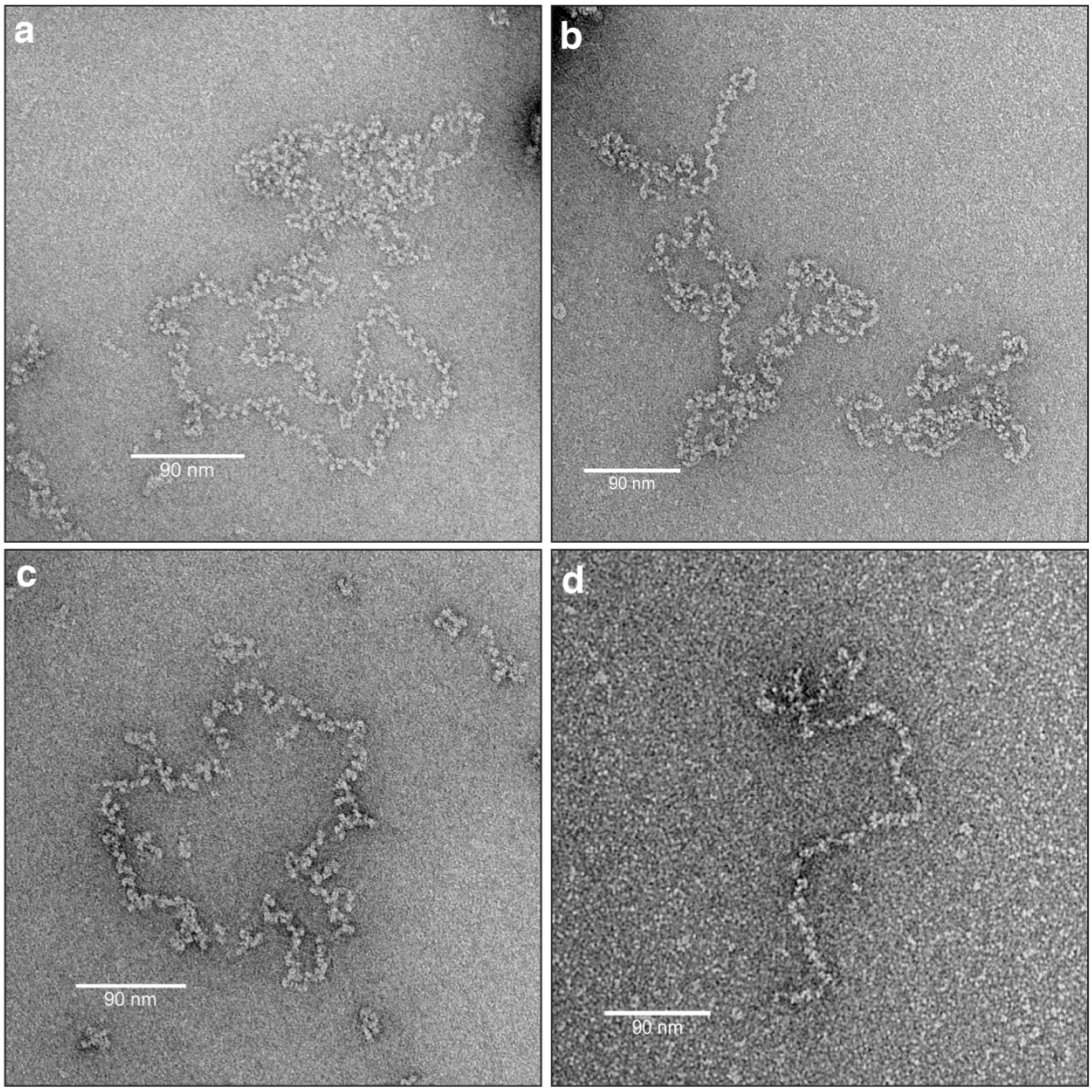


Figure 2.6 ***EM visualization of viral and reconstituted RNP*****.**
Authentic virus RNP showing the overall string-like appearance and lack of helical symmetry before (**a**) and after (**b**) overnight RNase treatment at room temperature. Reconstituted RNP with a 684-nt RNA before (**c**) and after (**d**) overnight RNase treatment at room temperature. The general appearance is like the authentic virus RNP in (**a**).

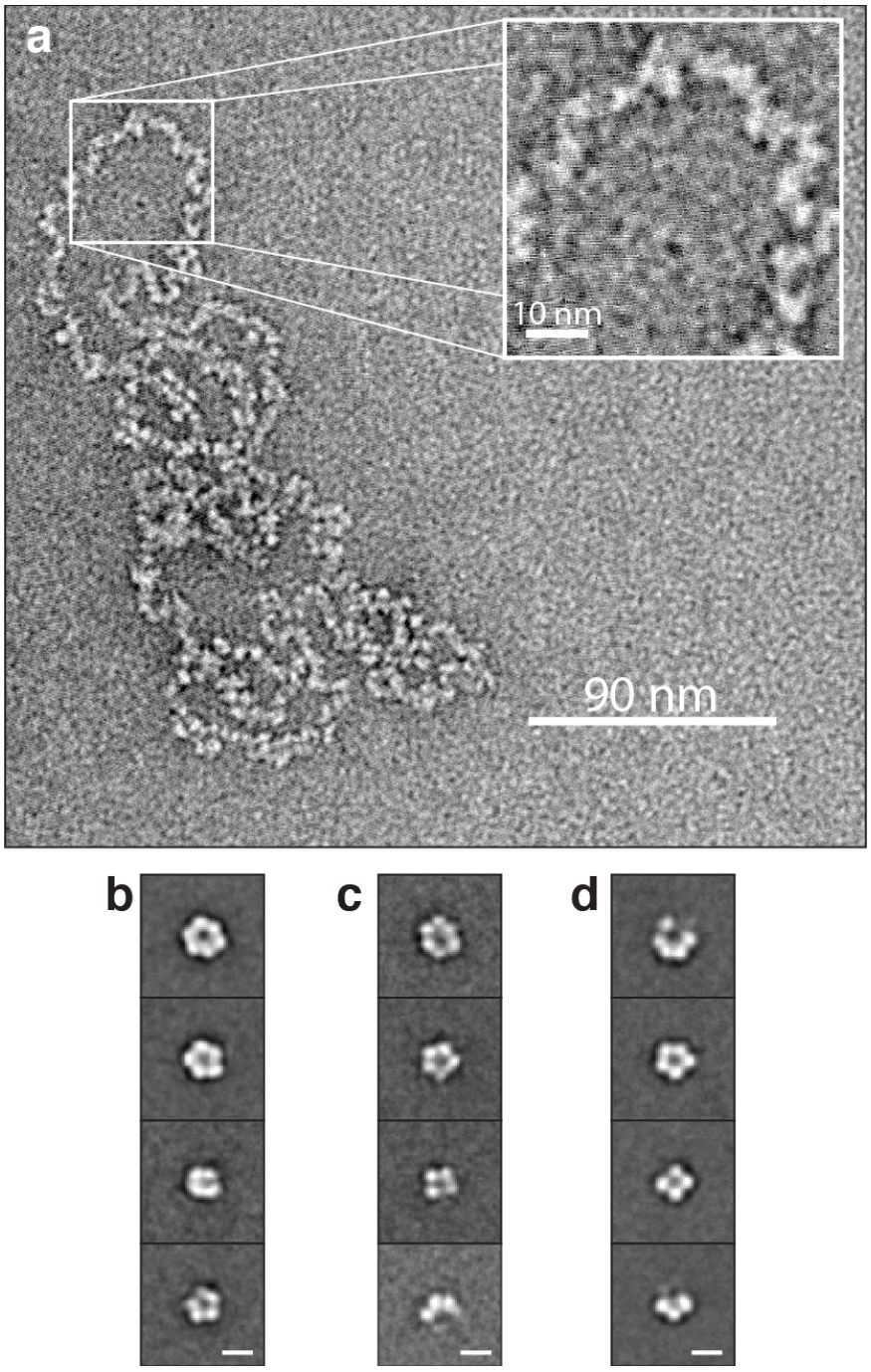


Figure 2.7 ***EM analysis of RNP*****.**
(**a**) Authentic virus RNP. Freshly prepared virus RNP has an overall string-like appearance and lacks helical symmetry. The inset shows an enlarged region. (**b**–**d**) Four representative class averages of virus (**b**), recombinant (**c**), and reconstituted (**d**) N-RNA multimers. (Scale bars, 11 nm)

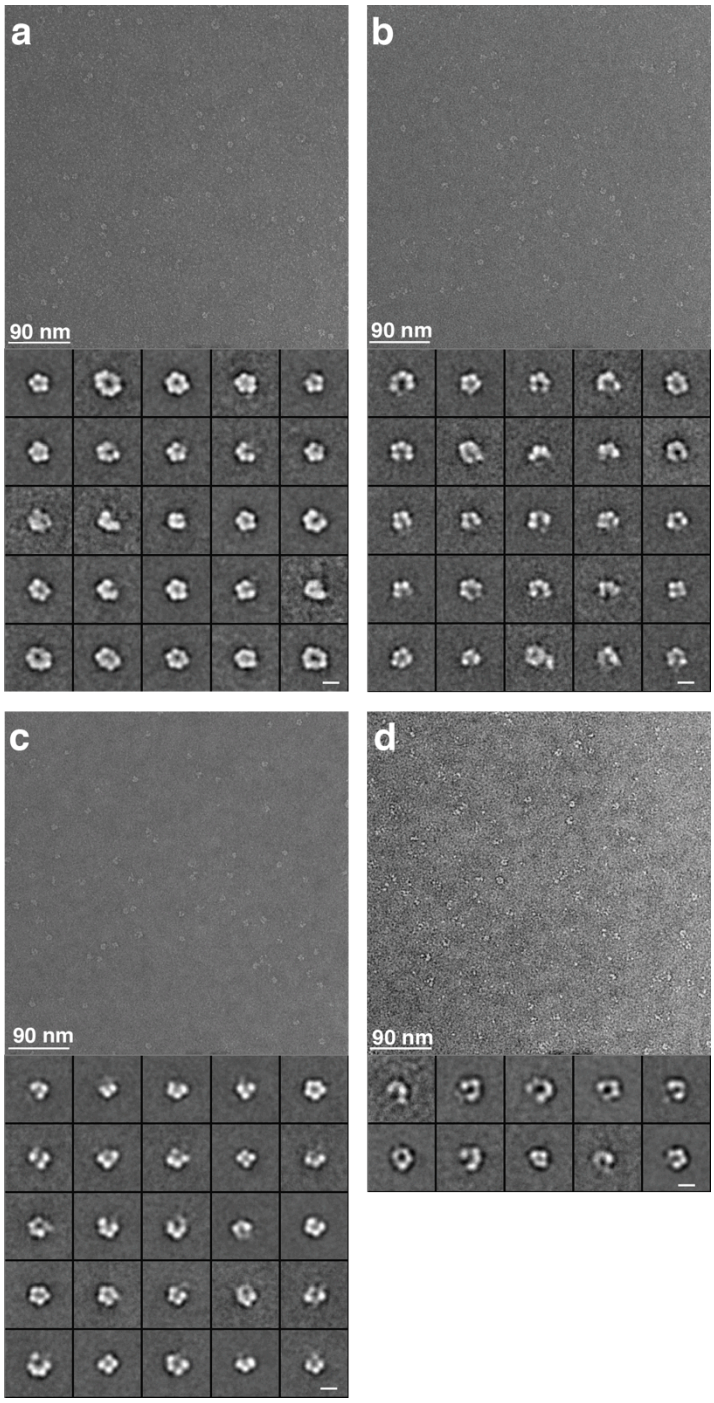


Figure 2.8 ***Single particle EM analysis of virus, recombinant, and reconstituted multimers.***
(**a**) N-RNA multimer from virus RNP and class averages from 1,019 particles. (**b**) N-RNA multimer from recombinant RNP and class averages from 782 particles. (**c**) N-RNA multimer reconstituted from N and RNA25 and class averages from 1,352 particles. (**d**) N-RNA multimer reconstituted from N and RNA10 and class averages from 411 particles. The upper panels are typical fields from which particles were picked. The lower panels show particles from 25 class averages, except for N-RNA10 with 10 class averages. (Scale bars, 11 nm.)

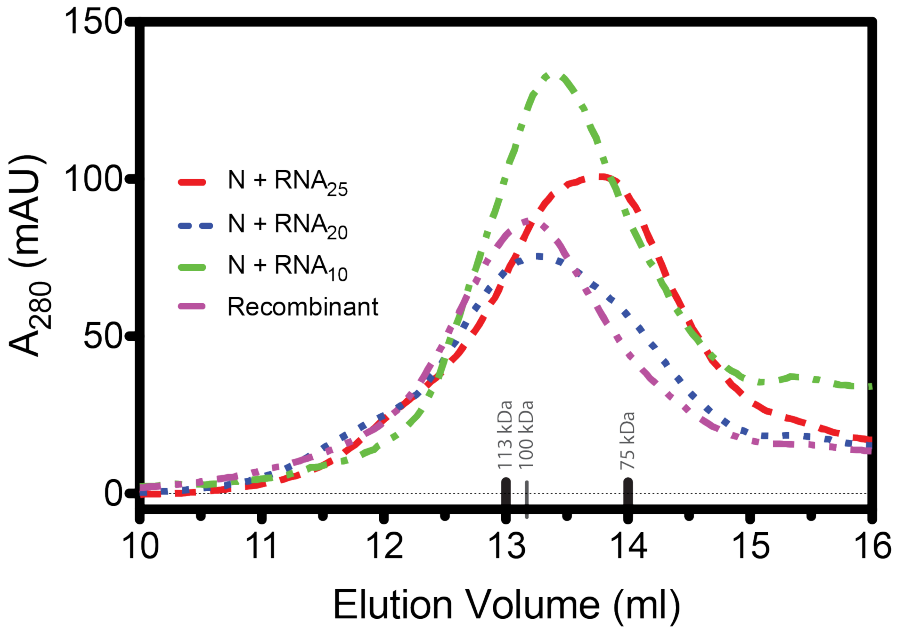


Figure 2.9 ***Gel-filtration profiles of N-RNA multimers.***
The elution volumes from an analytical S-200 column are similar for recombinant N-RNA purified from E. coli and reconstituted from RNA-free N and RNA10, RNA20, and RNA25. There was insufficient multimer released by ribonuclease treatment of virus RNP for gel-filtration analysis.

The stoichiometry of N and RNA in a reconstituted N-RNA multimer was estimated using known N and RNA extinction coefficients at 260 nm and 280 nm. We used a short RNA decamer in order to enhance the contribution of protein to the total absorbance, which was dominated by RNA. The reconstituted N-RNA_{10} multimer had similar behavior upon gel filtration (Fig. 2.9) and similar appearance in electron micrographs (Fig. 2.8) to the N-RNA_{25} multimer. The purified, reconstituted N-RNA_{10} multimer had an A_{260}/A_{280} ratio of 0.94, indicating an average stoichiometry of 5:1 N:RNA_{10}, in good agreement with the gel filtration data. This result is consistent with each bright globular density on multimer EM images corresponding to one N subunit.

Model quality and electron density

Recombinant, RNA-free RVFV N was crystallized and the structure was solved by multiwavelength anomalous diffraction from the selenomethionyl (SeMet) protein. The final model was determined at 1.93-Å with R/R_{free} of 0.211/0.254 (Table 1). MolProbity analysis of the refined structure gave good statistics with a MolProbity

score of 1.29 (99th percentile of 11840 structures in resolution range 1.93-Å ±0.25-Å) (Fig. 2.2)(119). Ramachandran analysis showed 99.6% of residues were in the allowed region, and 0.4% in the disallowed region. However, the electron density supported modeling the four residues in the disallowed regions. The crystals contained four N polypeptides in the asymmetric unit of space group *P*1, affording four independent views of the structure. The four copies of the N polypeptide are nearly identical with root-mean-square-deviations (RMSD) of 0.46 Å for 238 Cα atoms. The refined structure is complete with the exception of residues 16-19 and 28-30. In these regions, the electron density is not continuous and each of the four N polypeptides lacks density for one, four or seven amino acids.

Structure of N

RVFV N has a compact, helical fold consisting of N-terminal and C-terminal lobes of approximately equal size, connected by a linker helix (α7, residues 112-121) (Fig. 2.10a-b). Both the N-lobe (α1-α6, residues 1-111) and the C-lobe (α8-α12, residues 122-245) have a central helix (α3 and α9) surrounded by four or five other helices. Despite these similarities, the topologies differ and the N- and C-lobes cannot be superimposed. We examined the structural database for proteins with folds similar to RVFV N. Remarkably, the folds of both the N- and C-lobes appear to be novel. No structure similar to either lobe was identified in searches with the servers Dali (132) and VAST (133) servers.

The crystallized protein includes the full natural sequence (Met1-Ala245) without additional residues. Both chain termini are well ordered (Fig. 2.11). Met1 makes intra- and intermolecular contacts with hydrophobic residues in helix α1 (residues 3-10). The C-terminal α-carboxyl of Ala245 forms a salt bridge with the Arg178 side chain. Neither of the chain termini nor any loops protrude from the protein.

RVFV N crystallized as a symmetric dimer (Fig. 2.6c). This may be a natural dimer because the crystal contains two independent, nearly identical copies of the dimer (RSMD of 0.48 Å for 476 Cα atoms), and because a dimeric species was detected in solution (Figs 2.1b & 2.5). The dimer is formed by the contacts of residues in helices

α1, α7 and α8. The side chain of Trp125 (α7-α8 loop) is buried in the hydrophobic, water-free dimer interface where it contacts Met1, Gln5, Ile9 and Trp125 in the second monomer (Fig. 2.10d). Ala12, Val120, Val121, Glu124 and Thr131 also form inter-subunit van der Waals contacts. The small dimer interface (502 Å^2 buried surface area per monomer) is consistent with the low proportion of dimeric species in solution (Fig. 2.1b), nevertheless the dimer is expected to predominate at the high protein concentration in crystals.

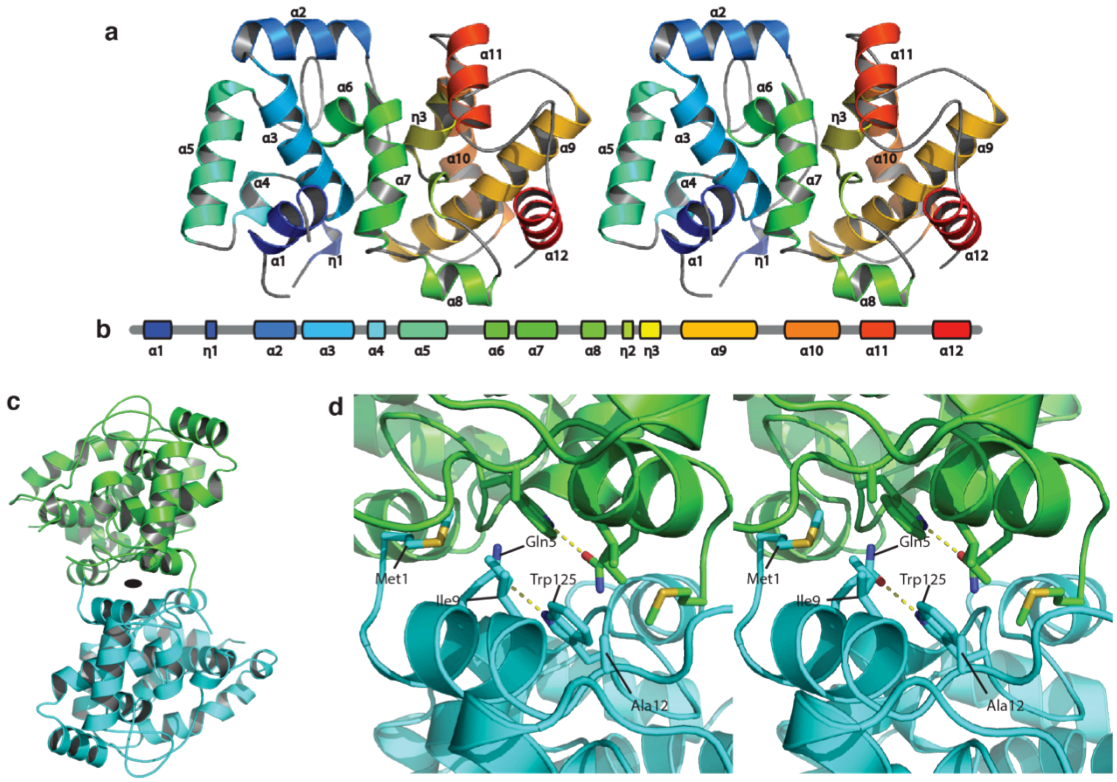


Figure 2.10 ***Structure of RVFV N.***
(**a**) Polypeptide fold. The stereo ribbon diagram is colored as a rainbow from blue at the N terminus to red at the C terminus with loops in gray. Helix α7, (vertical) in the center of the image, links the N lobe at the left and the C lobe at the right. (**b**) Diagram of helical secondary structure in the RVFN polypeptide. Colors are matched to A. (**c**) RVFV N dimer. In this view along the dimer axis, monomers are in green and cyan and the twofold axis is indicated by an ellipse. (**d**) Details of the dimer interface. The subunits are colored as in (**c**), side chains with dimer contacts are shown in stick form in the stereo view. Hydrogen bonds are shown as dashed lines.

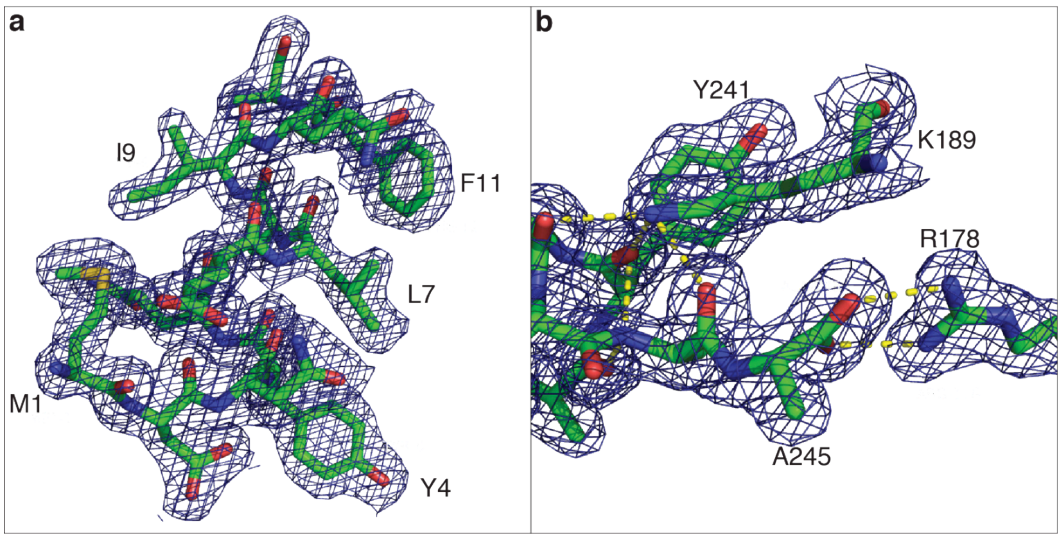


Figure 2.11 ***Electron density of RVFV N at the N and C termini.***
(**a**) $2F_o$-F_c map of the N-terminal helix contoured at 1σ. Residues 1–12 are shown in sticks. (**b**) $2F_o$-F_c map of residues 241–245 at the C terminus. Residues Arg178 and Lys189 are also depicted along with the salt-bridge and hydrogen bonds (dashed yellow lines) they form with the C terminus (Ala 245).

Comparison with N of Other Negative-Sense RNA Viruses

Given the rapid rate of virus divergence, we anticipated that the phlebovirus N might resemble the N of other negative-sense RNA viruses even though the sequences are dissimilar. Two different folds for N have been reported, one for FLUVA (92) in the family *Orthomyxoviridae*, and the other for four viruses in the order *Mononegavirales* (RABV (93), VSV (95), HRSV (94) and BDV (96)). However, the phlebovirus N fold differs from both these other N folds. Thus at least three different folds exist for N of the negative-sense RNA viruses. Intriguingly, all three folds are predominantly helical and are bi-lobed. However, the phlebovirus N has a more compact structure. RNA binds in a deep, positively charged cleft between the two lobes of N from both the *Mononegavirales* and FLUVA (93-97). Phlebovirus N lacks a cleft between the N- and C-lobes (Fig. 2.10). Another important difference is the lack of protrusions in phlebovirus N. The N- and C-termini of N of the *Mononegavirales* protrude from the subunit, as does an extended loop in the N of FLUVA. These protrusions contact other N subunits and are important to the structure of the RNP (93-97).

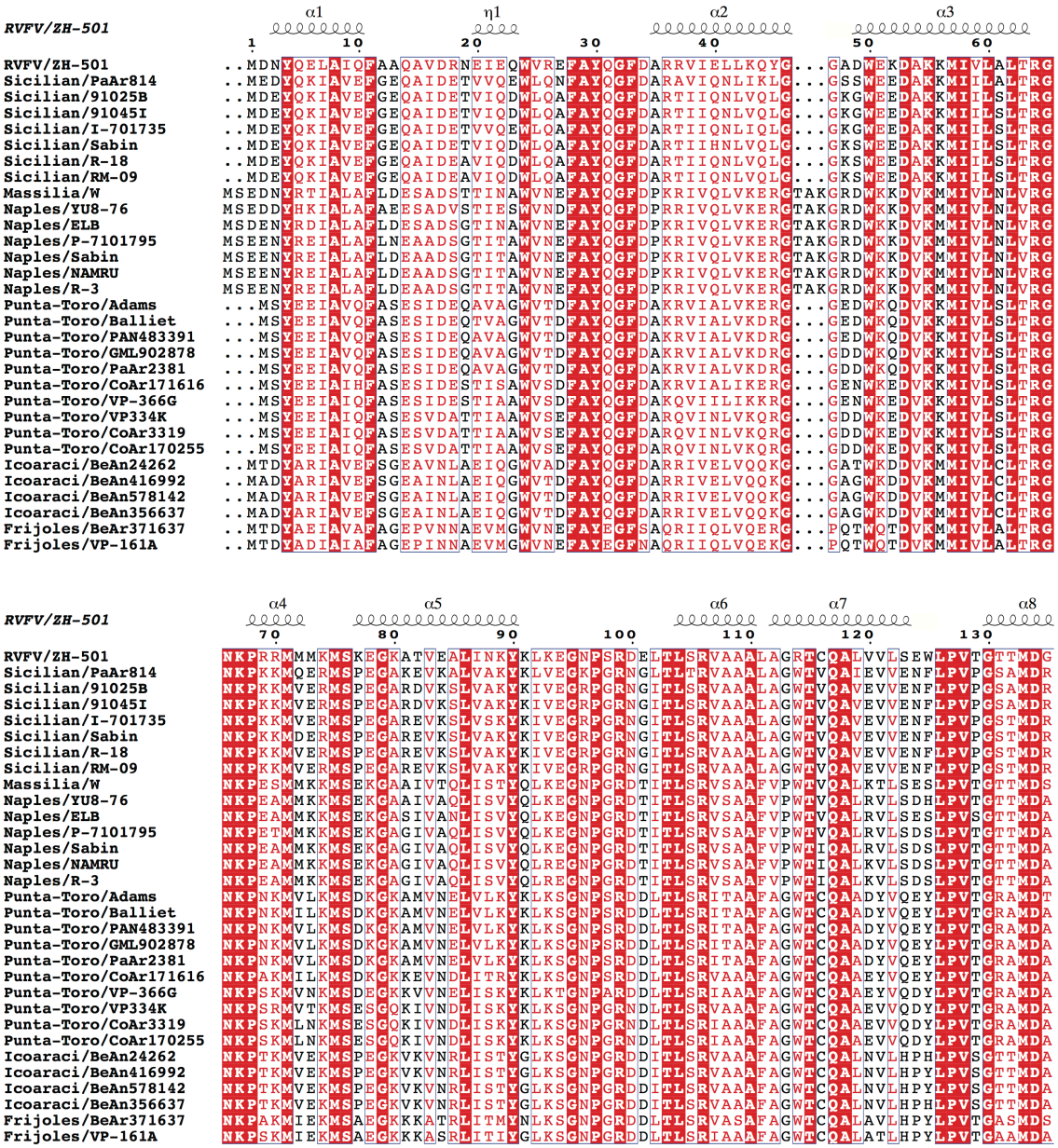
Conservation of Phlebovirus N

Among the highly conserved phlebovirus N (Fig. 2.12), a total 66 invariant residues map primarily to the core of the structure where they are important for conservation of the overall fold (Fig. 2.13a). Residues in the dimer interface are not strictly conserved, however a dimer contact appears feasible in all phlebovirus N because compensatory sequence changes accommodate the size and hydophobicity of the residue corresponding to Trp125. The structure is consistent with published mutagenesis data suggesting that the N-terminus of RVFV N is required for dimer formation (131). However, the conserved residues tested in the previous study (Tyr4, Phe11) point away from the dimer interface towards the inside of the monomer where they form stabilizing contacts in the hydrophobic core of the protein. The observed loss of dimer formation of Tyr4Gly and Phe11Gly (131) is likely due to destabilization of helix α1 (residues 3-10) and, indirectly, the dimer interface.

Mutagenesis of the Dimer Interface and C-Terminal Salt Bridge

The structure suggested that Trp125 is critical for dimer formation because of the hydrophobic contacts it makes with Met1, Ile9 and Trp125 of the opposing N monomer. Additionally, a salt bridge between the C-terminal carboxyl group of Ala245 and the Arg178 side chain, which is Arg or Lys in all phlebovirus N, may be important for overall structural integrity. To test the significance of these interactions, three mutant N alleles were generated, Trp125Ala, Arg178Gln and Arg178Glu, and the functionality of each was analyzed in a cell-based RdRp transcription assay in which RdRp and N were expressed from separate plasmids. When RdRp and N were both present and functional, a luciferase mRNA from a recombinant S segment was transcribed (126). The Trp125Ala mutant was severely compromised and activity was only 4% of the wild type allele (Table 2.5), suggesting that Trp125, and perhaps an N dimer, is essential for transcription. If the salt bridge of Arg178 with the C-terminus is critical, then the Arg178Gln mutant should retain more function than the Arg178Glu allele, and this was the observed result (Table

2.5). The activity of the Arg178Gln and Arg178Glu mutant was 25% and 7% of wild type, respectively. All alleles expressed protein at a level similar to wild type and all appeared capable of forming higher molecular weight complexes with RNA (Fig. 2.14, courtesy of Mary Piper).



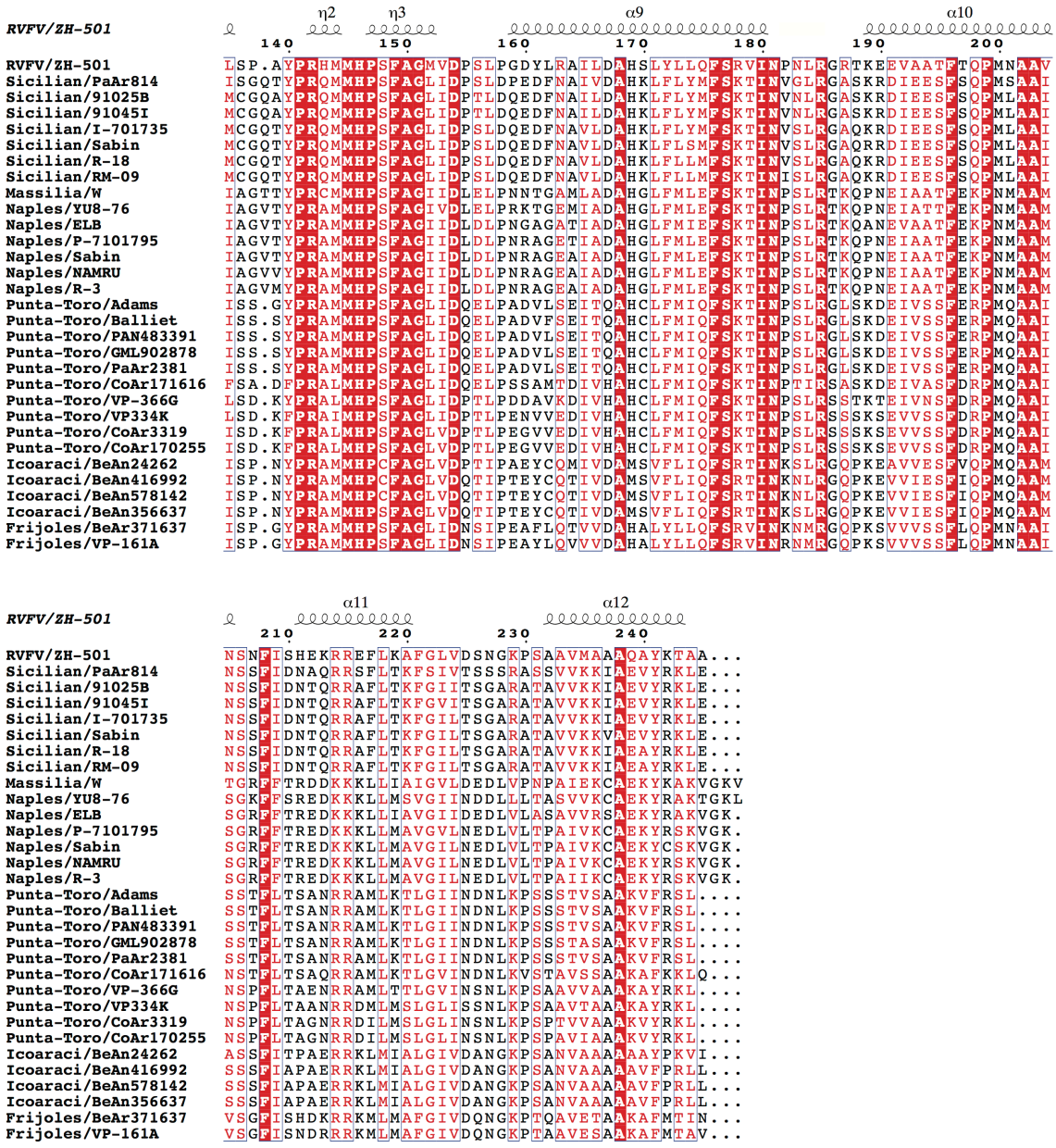


Figure 2.12 ***Sequence alignment of N from the phlebovirus genus.***
Invariant residues are show in white with red background, conserved residues are shown in red with white background, and variable residues are show in black with white background. The observed α and η (3_{10}) helices are indicated above the alignment. The sequence alignment was generated by clustalW (121) and the secondary structure annotations were assigned using ESPript (123).

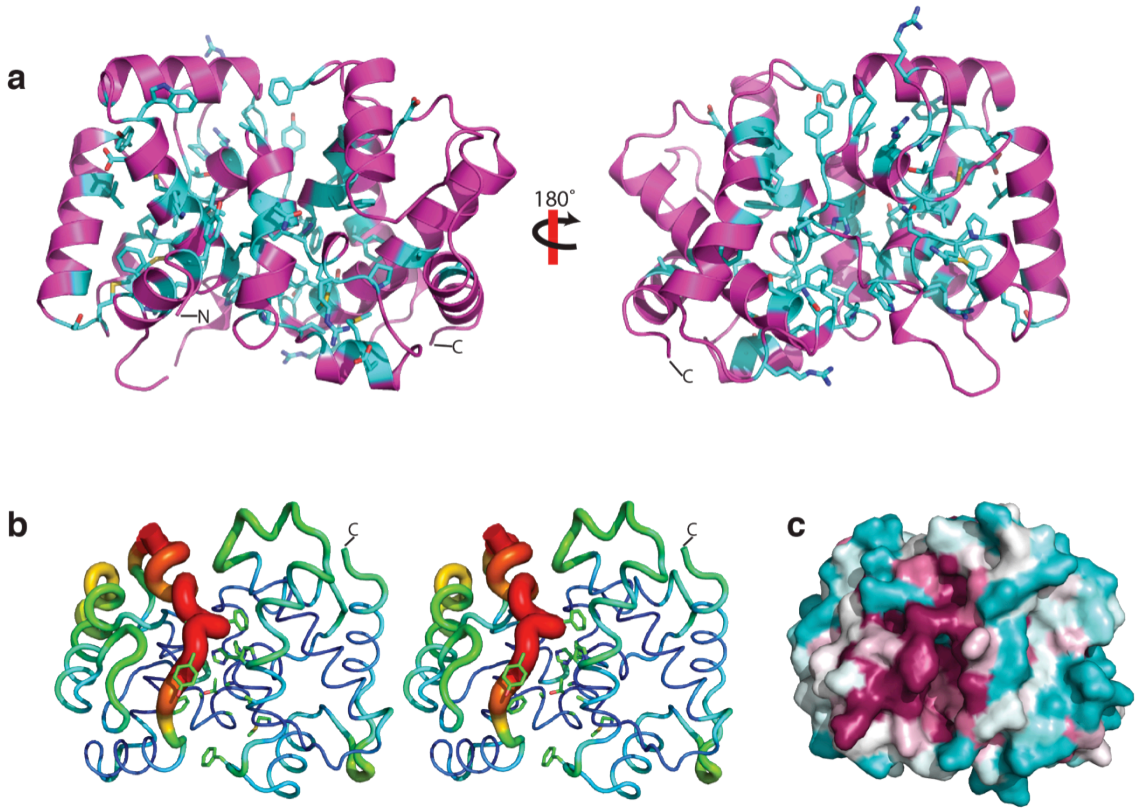


Figure 2.13 ***Conservation and hydrophobic surface pocket in RVFV N.***
(**a**) Sequence conservation mapped on the structure. In these front and back views of N, invariant residues (cyan, with side chains) are located primarily in the core of the protein. The front view here is identical to Fig. 2.10A. Invariance is defined according to Fig. 2.12. (**b**) Relative atomic mobility and the conserved hydrophobic pocket. N is viewed into the highly conserved hydrophobic pocket (view from the top of the left image in (**a**) above). The stereo image is colored as a rainbow according to average atomic B factors from low (blue, narrow tube) to high (red, wide tube). Side chains are shown for invariant (Phe28, Ala29, Tyr30, Phe33, Thr115, His146, Pro147, Ala150, Pro199, Ala202, Ala203, Phe208) and conserved (Ala109, Ala112, Met152) residues in the hydrophobic pocket.

Interaction Sites

We considered whether the RVFV N has an obvious RNA-binding surface. The N-lobe has a higher calculated isoelectric point (pI of 9.4 *vs*. 8.3) and a more positively charged surface (Fig. 2.15a) than the C-lobe. We compared RVFV N with structure-based homology models of N from other clades within the *Phlebovirus* genus. Based on the high sequence identity, N from all phleboviruses are expected to bind RNA similarly. There is a general trend of greater positive charge on the N-lobe than on the C-lobe, but the structures lack a common conserved basic surface. We also mapped sequence conservation onto the RVFV N surface (Fig. 2.15b). The most

strongly conserved surface is a hydrophobic pocket at the junction of the N- and C-lobes formed by a loop (residues 27-35) together with the C-terminal half of α10 and the five succeeding amino acids (residues 198-210) (Figs 2.13 & 2.15b). Residues 27-35 are among the most mobile regions of the N structure (Fig. 2.13b). The combination of mobility, conservation and hydrophobicity suggest that this site may be involved in a conserved protein-protein interaction.

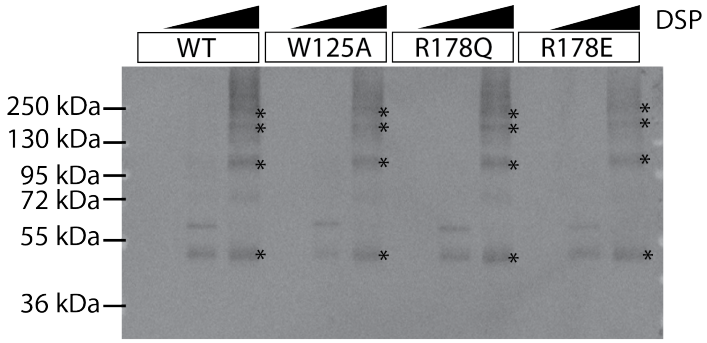


Figure 2.14 ***Protein levels and multimer formation for N mutants.***
The same cells used for the RdRp transcription assay (Table 2.5) were also analyzed for protein expression and N complex formation. Extracts were cross-linked using 0.0, 5.0 or 20.0 μM DSP. Protein complexes were then separated by SDS-PAGE and visualized by immunoblot. The dominant cross-linked species are indicated by asterisks and correspond to 2, 4 8 and 10 N. (Courtesy of Mary Piper)

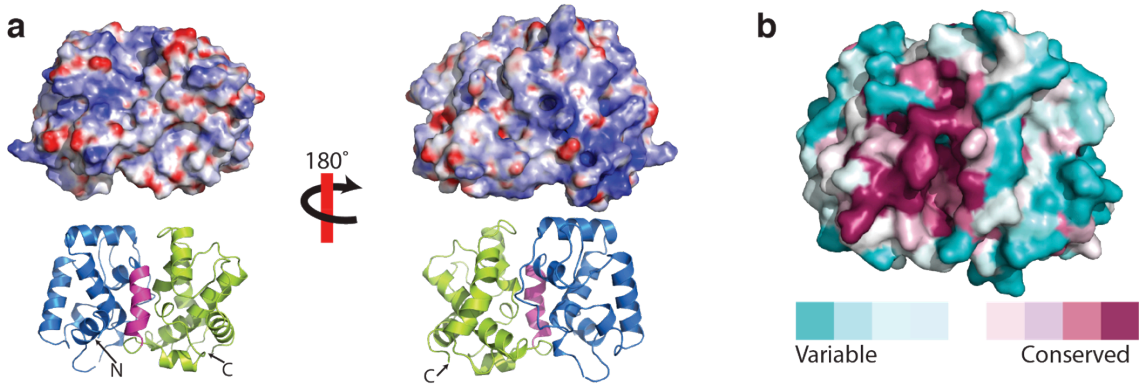


Figure 2.15 ***Properties of the RVFV N surface.***
(**a**) Electrostatic surface potential. The surface potential from -20 kT in red to +20 kT in blue is shown for the front and back of the RVFV N monomer. The ribbon diagrams below show the positions of the N-terminal (blue) and C-terminal (green) lobes and the linker (pink). The image at left is in the same orientation as (Fig. 2.10). (**b**) Conserved hydrophobic pocket. The most conserved surface of N is at the top relative to A. In this view, the N-terminal domain is at left and coloring is by conservation among phlebovirus N, as indicated.

Table 2.5 ***Effect of dimer and C-terminal salt-bridge substitutions on transcription.***

Allele	RLU (Std. dev.)	% Activity
WT	773 (169)	100
W125A	26.1 (2.8)	3.4
R178E	51.4 (9.4)	6.7
R178Q	193 (27)	25
No N	1.0 (0.1)	0.1

The wild type (WT), W125A, R178Q, and R178E N alleles were analyzed for function by the transcription assay described in the Material and Methods. Renilla luciferase (RLU) activity was measured at 48 h post-infection and is expressed relative to the no-N control and as a percent of WT activity. Data are the average of six experiments, with standard deviations in parentheses.

Discussion

The structure of RVFV N is the first for a nucleocapsid protein from any virus in the *Bunyaviridae* family. The novel protein fold is an addition to nature's repertoire of RNA-binding proteins (Fig. 2.10). High levels of sequence identity (36-59%) assure that all phlebovirus N possess a similar fold, and also suggests that all phlebovirus N bind RNA similarly.

This work establishes that RNP organization in the *Phleboviruses* is different than in other negative-sense RNA viruses. Structures have been reported for RNPs from four negative-sense RNA viruses (*Mononegavirales* and *Orthomyxoviruses)* (83, 93-95). In all cases, RNA binds nonspecifically in an electropositive cleft between the lobes of the N subunit. RNP oligomers from these viruses have a similar architecture in which RNA binds around either the outside or inside of a ring of 9-11 N subunits. In all cases, protrusions from the N subunits make specific contacts with adjacent subunits to maintain the ring structure. In some cases, the number of subunits in the oligomer ring matches the helical repeat of the polymeric RNP, which is apparent by EM. For HRSV, each N subunit also interacts with other N subunits in the preceding or following turns of the helical nucleocapsid (94).

In contrast to the RNPs of *Mononegavirales* and *Orthomyxoviruses*, no helical structure was apparent in any electron micrographs of RVFV authentic or reconstituted RNP (Figs 2.6 & 2.7a). Our results are similar to early electron micrographs of bunyavirus RNP, which lacked helical symmetry and also suggested that bunyavirus RNPs form large macro-circles (99), probably due to pairing of 10-15 complementary bases at the 3' and 5' ends of each genomic segment (134). RVFV RNP macro-circles were seen in some of our images. Aggressive ribonuclease treatment of RVFV RNP released an N-RNA multimer that appears heterogeneous in composition and has an average diameter of 11 nm (Figs 2.7b & 2.8a). Multimers of nearly identical appearance to those from virus RNP were observed both for ribonuclease-treated recombinant RNP and for RNP reconstituted with small RNAs (Figs 2.7c-d & 2.8). The heterogeneity of these minimal particles was surprising, but

is consistent with the lack of helical symmetry in the RNP. The N crystal structure is also consistent with the lack of helical symmetry. The highly compact phlebovirus N has no protruding loops or termini that could link it to other N molecules in a super-structure like the rings of 9-11 subunits observed for the *Mononegavirales* and *Orthomyxoviruses* (93-95, 97), although we cannot rule out the possibility of major conformational changes to N upon RNA binding. We observed no large super-structure for recombinant RVFV N in solution, unlike the recombinant rings purified for N from RABV (93), HRSV (94), VSV (95) and FLUVA (97). All N-RNA multimers that we could test by gel filtration had a similar apparent molecular weight of ~100 kDa (Fig. 2.9).

Our working model for the structure of phlebovirus RNP is that N binds cooperatively to RNA. However, the cooperativity is limited to 4-7 N subunits based on the similar size and appearance of multimers from viral, recombinant and reconstituted RNP and their similar behavior upon gel filtration (Figs 2.7, 2.8 & 2.9). The RNP lacks a strong helical structure, and N-N contacts are too weak for ribonuclease treatment to release a specific N-RNA oligomer from virus RNP. Instead of a tightly associated, symmetric N-RNA oligomer, weak protein-protein interactions lead to a mixture of multimer species. If there is a weakly associated fundamental oligomer, it may be a hexamer because our cross-linking data showed a preponderance of species containing multiples of two N subunits (Fig. 2.5). Overall, the organization of phlebovirus RNP appears less symmetric and with few specific protein-protein interactions compared to the helical RNP of the *Mononegavirales* and *Orthomyxoviruses*. The evolutionary path to phlebovirus RNP is not common with the path to helical RNP of some other negative-sense RNA viruses.

The unusual N-RNA multimer and perhaps the non-helical RNP may be a general property of the *Bunyaviridae*. The observed RVFV N-RNA multimer species are similar to the reported 109-kDa recombinant RNP from Bunyamwera virus (85). Bunyamwera virus and RVFV belong to different genera within family *Bunyaviridae* and their N are not obviously similar at the amino acid level.

The N-lobe of RVFV N was identified as a potential RNA interaction site because it is more positively charged than the C-lobe in N from all phlebovirus clades. In whatever manner N binds RNA, it is expected to engage the phosphate backbone because the multimer is so highly ribonuclease resistant. EM visualization does not permit assignment of RNA to a specific location in the multimer particles.

The relevance of the N dimer to the RNP is unclear. Based on the size of multimers in electron micrographs (Figs 2.7 & 2.8) and their gel filtration profile (Fig. 2.9), each bright globular density in the multimer images must be a monomer and not a dimer of N. A precedent for different oligomer species for RNA-free and RNA-bound N exists for FLUA where the free protein is a trimer and the RNP is a nonamer (92, 97). We probed the RVFV N dimer interface by site-directed mutagenesis at Trp125, but found cross-linked RNA complexes similar to the wild type (Fig. 2.14). However, the Ala substitution revealed an important role for Trp125 in replication (82). This could be due to any number of molecular interactions, most obviously with the RdRp.

The most highly conserved surface of phlebovirus N is a hydrophobic pocket at the interface of the N- and C-terminal lobes (Figs 2.13 & 2.15,). The conservation in this region suggests an important function that is common to phleboviruses, and the hydrophobicity of the surface suggests that it is not a site for RNA binding but rather is for interaction with a viral or (unidentified) host protein. Among potential viral protein partners, the RdRp is an obvious possibility because N is required for transcription and replication by the RdRp (84). However, several lines of evidence suggest that an envelope glycoprotein may be the target of the conserved hydrophobic pocket on N.

Packaging of RNPs into virions occurs at a site of virus assembly on the Golgi membrane (135). The cytoplasmic tail of the RVFV envelope glycoprotein G_N was shown recently to recruit the encapsidated genome to the Golgi membrane prior to virion assembly (136). Moreover, three regions within the G_N tail of the Uukuniemi virus were shown to be important for nucleoprotein binding to the glycoproteins (32). Genome recruitment is expected to be similar in all phleboviruses, thus the

conserved hydrophobic pocket of N is a candidate G_N binding site. This hypothesis is consistent with the ability of bunyaviruses, both in nature and *in vitro*, to undergo reassortment in which progeny have genomic segments that derive from more than one parental virus (137-139). Reassortment requires promiscuity in the interaction of N with genomic RNAs from heterologous viruses and in protein-protein interactions necessary for assembling virions. All characterized reassortant bunyaviruses isolated in nature are M segment reassortants (140, 141), demonstrating that the envelope glycoproteins, which are encoded by the M segment, are capable of interacting with heterologous RNPs. The hydrophobic character of some regions of the G_N-tail, as well as Pro and Trp residues within it, are conserved amongst phleboviruses (105) and could function in protein-protein interactions with N. Whether the conserved pocket of N interacts with the G_N cytoplasmic tail, with the RdRp or with a host protein, it has potential as a drug target because it is conserved in phleboviruses.

The structure and characterization of *Phlebovirus* N and RNP reveal a new paradigm for encapsidation of the genomes of negative-sense RNA viruses, provide a platform for further studies of virus pathogenicity, and suggest a potential site for development of effective antiviral therapeutics.

Addendum to thesis

The data presented in Chapter 3 revealed that the RVFV N monomeric structure is the closed conformation, which prevents N multimerizing and nucleic acid binding. Therefore, the conclusions drawn from the monomeric structure have changed.

The N dimer interface observed in the crystal structure is not physiologically relevant. It is a concentration-dependent multimeric state of N that occurs in the absence of RNA. The conserved hydrophobic pocket is involved in N multimerization. The positive surfaces at the N-lobe are not involved in N multimeriation or RNA binding. Trp125 plays a role in N multimerization and Arg178 forms a salt bridge with the C-terminal carboxylic acid to create the compact core structure. Please see Chapter 3 for more detailed information.

CHAPTER 3

STRUCTURES OF PHLEBOVIRUS RIBONUCLEOPROTEINS

Summary

Rift Valley fever virus (RVFV) and Toscana virus (TOSV) are human pathogens for which no effective therapeutics exist. The membrane envelopes of these arthropod-borne viruses from the *Phleboviruses* genus (*Bunyaviridae* family) enclose segmented negative-sense RNA genomes that are bound by nucleocapsid protein (N) to form a ribonucleoprotein (RNP) complex. The interactions between the N and genomic RNA required to form the RNPs have not been characterized at high resolution. Crystal structures of RNA-free N from RVFV elucidated the core structure of N, but provided no details about RNA binding, the formation of the asymmetric RNP, or the relationship of the asymmetric RNP to the multimers released by ribonuclease digestion. In this study, we present crystal structures of phlebovirus N-RNA complexes reconstituted with defined RNAs of different lengths. These first high-resolution views reveal a surprising diversity of binding geometries where variation in the RNA length and the number of nucleotides per N subunit give rise to tetrameric, pentameric and hexameric structures. RVFV N binds RNA by sequestering all nucleotide bases in an RNA-binding slot with the sugar-phosphate backbone facing the solvent. The structures reveal multiple conformations of an N-terminal helical arm. The flexibility of the arm allows for the asymmetry of the N-RNA multimers and the asymmetric architecture of the RNP. Despite the lack of symmetry in all multimers, the crystal structures reveal a common building block consisting of the core domain of an N subunit, four RNA nucleotides, and the helical arm of an adjacent subunit (RNA-N_{core}-arm). Together with direct binding measurements showing that N binds nucleic acids with equal affinity regardless of length or sequence, these results provide a model for sequence-independent base sequestration by N and explain the observed structure of phlebovirus RNP.

Publication note: Research described in Chapter 3 has been submitted for publication in *PNAS*.

Introduction

Viruses of the *Phlebovirus* genus (*Bunyaviridae* family) are transmitted by arthropod vectors and cause a variety of severe diseases worldwide. The Rift Valley fever virus is a highly infectious, mosquito-borne pathogen endemic to sub-Saharan Africa. RVFV infects livestock and humans, and generally causes a flu-like illness; however, 1% of cases result in hemorrhagic fever disease, which has a 50% case-fatality rate (142). The closely related Toscana virus (TOSV) is endemic to the Mediterranean basin, is transmitted by infected phlebotomine sandflies and causes neurological dysfunction in humans (143). The membrane envelope of these viruses encloses a three-segment, negative-sense RNA genome that is encapsidated by a nucleocapsid protein (N), forming the ribonucleoprotein (RNP) (84).

Encapsidated genomes are required for RNA-dependent RNA polymerase (RdRp) function because the polymerase cannot replicate or transcribe naked RNA (31, 144). High-resolution crystal structures are available for RNP from three viruses in the *Rhabdo-* and *Paramyxoviridae* families (93-95). Each structure showed protrusions from the N core interacting with neighboring subunits to form a helical N-RNA polymer. N subunits spontaneously form multimers to encapsidate viral RNA in a ribonuclease-resistant RNP that prevents RNA base pairing and inhibits the antiviral response in infected cells. This mode of N-N assembly allows N an unlimited capacity for multimerization restricted only by the length of RNA being encapsidated. A positively charged RNA-binding surface in N becomes a continuous groove upon N multimerization and allows single-stranded RNA to thread through the growing RNP (93-95).

Phlebovirus N are among the smallest N of negative-sense RNA viruses (NSV), yet are required for multiple functions during the viral replicative cycle, including the replication and transcription of the genome (84). Understanding the mechanism of interaction between the RNA genome and N in phleboviruses would elucidate their genome-packaging strategy and may allow for the targeted development of therapeutics to treat infections.

We showed previously that phleboviruses have a unique genome packaging strategy, and an RNP that lacks the helical symmetry observed in some other negative-sense RNA viruses (NSVs) (33, 145). However, the detailed interactions between phlebovirus N and the viral RNA genome are unknown. The crystal structure of a RVFV N monomer revealed a compact structure of novel fold (145). The structure of an RVFV N hexamer demonstrated conformational flexibility in N and showed a putative RNA binding site on the inner surface of the hexameric ring (146). An α-helical arm, which is sequestered within the subunit in the N monomer structure, extends from the monomer to mediate subunit contacts in the hexamer. However, both structures lack RNA and a detailed explanation for the non-helical structure of the N-RNA polymer has not been provided.

EM visualization of authentic RNPs from phlebovirus-infected cells revealed an extended, open RNP that lacks higher-order structure or symmetry (Fig. 2.7) (99, 145). Nucleocapsid protein-RNA (N-RNA) complexes extracted from viral RNPs by extensive ribonuclease treatment or expressed recombinantly have asymmetric ring-like structures of variable size (145, 146). Single-particle EM analysis suggested a heterogeneous population of multimers, each with 3-7 N subunits (Figs 2.7 & 8). The heterogeneous, recombinant N-RNA multimers did not crystallize even after extensive ribonuclease digestion and purification.

In this study we used fluorescence polarization to investigate the N-genome interaction using RNA-free N and defined RNA and DNA oligomers. We solved structures of apo-N hexameric rings from RVFV and TOSV. The hexameric structure showed N has an identical core structure to the N monomer except for the extension of a helical arm. The helical arm unfolds from the core of N to contact neighboring subunits. We used electron microscopy to screen nucleic acid lengths for the most homogeneous reconstituted N-DNA complexes for crystallization trials. Homogeneous N complexes allowed us to solve crystal structures of reconstituted N-RNA and N-DNA complexes. The new crystal structures show the tremendous flexibility of the helical arm, which allows phlebovirus N to form several distinct multimers. The N-RNA structures also revealed a RNA-binding slot, where nucleic

acid bases are non-specifically sequestered from solvent. The new structures provide exquisite details about N organization and RNA binding and explain the asymmetry in phlebovirus RNPs.

Experimental Procedures

Plasmid Construction

All plasmids were generated using standard molecular cloning techniques and confirmed by sequencing. The plasmid encoding Rift Valley fever virus (RVFV) N (pIPER1) was produced by Mary Piper and is described in chapter 2 (145). To construct the pTOSVN plasmid, the full-length Toscana virus (TOSV) gene encoding N was amplified using primers 5′agattggtggcATGTCAGACGAGAATTATCGC and 5′gaggagagtttagacattaCTTGCCAACCTTGGCG (IDT). The PCR product was processed with T4 DNA polymerase (Promega) and dGTP (Invitrogen) to generate an overhang for ligation-independent cloning (LIC) and then incubated with processed LIC-SUMO vector (pETHSUL), which encodes a His_6-SUMO fusion protein and was provided by P. Loll, Drexel University (147). The mixture was transformed into *E. coli* XL1-Blue competent cells for ligation and amplification of the plasmid.

Productions and Purification of Recombinant Proteins

Recombinant RVFV N and TOSV N were purified under native or denaturing conditions, as described previously (145). Briefly, recombinant proteins were purified by Ni-affinity chromatography, the His_6-SUMO tag was removed by SUMO protease, and proteins were further purified by gel filtration. RNA-free proteins were generated by Ni-affinity purification, denaturation, refolding, tag cleavage, and gel filtration. Proteins were stored at -80˚C in 20 mM Tris pH 7.8, 500 mM NaCl and 10% glycerol.

Fluorescence Polarization Binding Experiments

Fluorescence polarization (FP) experiments were performed using native or refolded N proteins and PAGE-purified, 5'-6-FAM™ fluorescein-labeled single-stranded RNA or DNA oligomers (IDT). Prior to FP experiments, the N proteins were dialyzed against FP buffer (20 mM Tris pH 7.8, 50 mM NaCl). FP measurements were performed on a SpectraMax M5 (Molecular Devices) using 96-well black microplates (Fluotrac 200 by Greiner Bio-One) and a reaction mix containing 10 mM

Tris pH 7.8, 25 mM NaCl, 5 nM RNA or DNA. N was serially diluted 10-fold into FP buffer, and 45 µL of each dilution was added to 45 µL of 10 nM labeled RNA or DNA. All experiments were done in duplicate with appropriate controls. Binding constants (K_d) were determined by fitting FP data by nonlinear least-squares regression to FP = B_{max} ×[N]/(K_d +[N]) where B_{max} is the FP signal at maximum binding in KaleidaGraph (version 4.1, Synergy Software, Reading, PA, USA) (148). RNA binding to the RVFV recombinant N multimer did not reach saturation, so B_{max} was approximated as the average millipolarization (mP) values for fully bound RNA-free N and recombinant N monomer.

Electron microscopy

Negative stained samples for electron microscopy were prepared as described (145). Imaging was performed at room temperature with a Tecnai T12 transmission electron microscope operated at an acceleration voltage of 120 kV. Images of RVFV N-DNA_{35} and N-DNA_{40} were recorded on a mounted Gatan US4000 CCD camera at a magnification of 71,138X and a defocus value of ~1.5 µm. All images were binned (2 x 2 pixels) to obtain a pixel size of 4.16 Å on the specimen level. For 2D reference-free alignment and classification of particles, 3770 particles of RVFV N-DNA_{35} were selected from micrographs using EMAN Boxer (129). Reference-free alignment, classification into ninety-nine groups, and class averaging were performed with EMAN (129).

Crystallization

RVFV N-RNA_{28}, N-RNA_{35} and N-DNA_{30} complexes were produced by 1-hr incubation of RNA-free N with PAGE-purified nucleic acid at a ratio of 8 N to 1 RNA poly-uracil oligomer or DNA poly-thymidine oligomer. Complexes were separated from excess N by gel filtration. Prior to crystallization, all samples were dialyzed against crystallization buffer (20 mM Tris pH 7.8, 250 mM NaCl) and maintained at a concentration of ~10 mg/mL. All crystals were grown by hanging-drop vapor diffusion, cryo-protected by a 10 second soak in well solution augmented with cryo-protectant, harvested into loops, and flash-cooled by plunging into liquid N_2.

RVFV N_4-RNA_{28} crystallized at 20˚C in ~7 days from a 1:1 mixture of N-RNA_{28} and well solution containing 18% PEG 3350, 350 mM NaCl, 100 mM Bis-tris pH 5.5. RVFV N_5-RNA_{35} crystallized at 20˚C in ~4 weeks from a 1:1 mixture of N-RNA_{35} and well solution containing 28% PEG 3350, 280 mM $(NH_4)_2SO_4$. RVFV N_6-RNA_{35} crystallized at 20˚C in ~4 weeks from a 1:1 mixture of N-RNA_{35} and well solution containing 24% PEG 3350, 350 mM $(NH_4)_2SO_4$. RVFV N-DNA_{30} crystallized at 20˚C in ~4 days from a 1:1 mixture of N-DNA_{30} and well solution containing 15% PEG 3350, 350 mM NaCl, 100 mM Bis-tris pH 5.5. RVFV N crystallized at 20˚C in ~7 days from a 1:1 mixture of protein and well solution containing 15% PEG 3350, 150 mM $(NH_4)_2SO_4$. Recombinant TOSV N crystallized at 4˚C in ~7 days from a 1:1 mixture of protein and well solution containing 23% isopropanol, 280 mM $MgCl_2$, 100 mM Hepes pH 7.5. Cryo-protectants were 10% glycerol for RVFV N_5-RNA_{35}, RVFV N_6-RNA_{35}, RVFV N_6-DNA_{30} and RVFV N_6; 15% glycerol for RVFV N_4-RNA_{28}; and 30% MPD for TOSV N_6.

Data Collection and Structure Determination

Diffraction data were collected at 100K on GM/CA beamlines 23ID-B (RVFV N_4-RNA_{28}, N_5-RNA_{35}, N_6-RNA_{35} and N_6) and 23ID-D (RVFV N_6-DNA_{30} and TOSV N_6) at the Advanced Photon Source, Argonne National Laboratory (Argonne, IL) at λ = 1.0332 Å. Diffraction images were indexed and integrated using iMOSFLM (110) and data were scaled using SCALA (Tables 2.1-7) (112), both part of the CCP4 suite (118). A truncated version of monomeric RVFV N (3LYF (145), residues 35-245) was used as a probe for molecular replacement with PHASER (149). Modeling was completed manually using Coot (116) and RCrane (150). Refinement was performed using BUSTER (151) with non-crystallographic symmetry (NCS) restraints and translation-libration-screw (TLS) parameterization of molecular motion (Table 2.7) (152). Individual isotropic thermal parameters were refined for structures with d_{min} less than 3.0 Å, and grouped isotropic thermal parameters for those with d_{min} greater than 3.0 Å. Ramachandran analysis and structure validation were performed by MolProbity (119). Ramachandran plots (Figs 3.1-3.6) of each structure showed that at least 98% of residues were in the allowed region: RVFV N_4-RNA_{28}, 99.4%;

RVFV N_5-RNA_{35}, 98.9%; RVFV N_6-RNA_{35}, 98.8%; RVFV N_6-DNA_{30}, 99.6%; RVFV N_6, 99.1%; TOSV N_6, 99.5%. PyMOL was used for structural alignments and to generate figures (120). The APBS plugin in PyMOL was used to calculate electrostatic surface potentials (122).

Table 3.1 ***SCALA scaling summary for RVFV N_4-RNA_{28}.***

	Overall	InnerShell	OuterShell
Low resolution limit	48.42	48.42	2.27
High resolution limit	2.15	6.80	2.15
Rmerge	0.111	0.045	0.584
Rmerge in top intensity bin	0.043	–	–
Rmeas (within I+/I–)	0.130	0.053	0.696
Rmeas (all I+ & I–)	0.130	0.053	0.696
Rpim (within I+/I–)	0.067	0.028	0.375
Rpim (all I+ & I–)	0.067	0.028	0.375
Fractional partial bias	–0.035	–0.047	–0.132
Total number of observations	206115	7092	21896
Total number unique	55733	1869	6638
Mean((I)/sd(I))	9.3	24.2	2.3
Completeness	95.5	99.1	77.9
Multiplicity	3.7	3.8	3.3

Table 3.2 ***SCALA scaling summary for RVFV N_5-RNA_{35}.***

	Overall	InnerShell	OuterShell
Low resolution limit	42.38	42.38	4.11
High resolution limit	3.90	12.33	3.90
Rmerge	0.212	0.210	0.463
Rmerge in top intensity bin	0.178	–	–
Rmeas (within I+/I–)	0.300	0.297	0.655
Rmeas (all I+ & I–)	0.300	0.297	0.655
Rpim (within I+/I–)	0.212	0.210	0.463
Rpim (all I+ & I–)	0.212	0.210	0.463
Fractional partial bias	–0.735	–0.964	–0.637
Total number of observations	57521	1640	8356
Total number unique	28961	866	4197
Mean((I)/sd(I))	2.8	4.6	1.6
Completeness	99.1	93.9	99.1
Multiplicity	2.0	1.9	2.0

Table 3.3 ***SCALA scaling summary for RVFV N_6-RNA_{35}.***

	Overall	InnerShell	OuterShell
Low resolution limit	57.42	57.42	3.58
High resolution limit	3.40	10.75	3.40
Rmerge	0.190	0.050	0.513
Rmerge in top intensity bin	0.048	–	–
Rmeas (within I+/I–)	0.268	0.070	0.725
Rmeas (all I+ & I–)	0.268	0.070	0.725
Rpim (within I+/I–)	0.190	0.050	0.513
Rpim (all I+ & I–)	0.190	0.050	0.513
Fractional partial bias	–0.069	–0.065	–0.172
Total number of observations	224721	7343	31220
Total number unique	117933	3769	16752
Mean((I)/sd(I))	4.0	9.6	1.7
Completeness	94.7	95.6	92.3
Multiplicity	1.9	1.9	1.9

Table 3.4 ***SCALA scaling summary for RVFV N_6-DNA_{30}.***

	Overall	InnerShell	OuterShell
Low resolution limit	53.66	53.66	2.85
High resolution limit	2.70	8.54	2.70
Rmerge	0.114	0.031	0.779
Rmerge in top intensity bin	0.037	–	–
Rmeas (within I+/I–)	0.129	0.036	0.887
Rmeas (all I+ & I–)	0.129	0.036	0.887
Rpim (within I+/I–)	0.060	0.017	0.418
Rpim (all I+ & I–)	0.060	0.017	0.418
Fractional partial bias	–0.049	–0.064	–0.161
Total number of observations	112943	3587	15490
Total number unique	25771	902	3615
Mean((I)/sd(I))	10.0	26.3	2.0
Completeness	99.6	95.5	98.5
Multiplicity	4.4	4.0	4.3

Table 3.5 ***SCALA scaling summary for RVFV N_6.***

	Overall	InnerShell	OuterShell
Low resolution limit	52.45	52.45	3.27
High resolution limit	3.10	9.80	3.10
Rmerge	0.098	0.025	0.693
Rmerge in top intensity bin	0.025	–	–
Rmeas (within I+/I–)	0.107	0.027	0.768
Rmeas (all I+ & I–)	0.107	0.027	0.768
Rpim (within I+/I–)	0.041	0.011	0.320
Rpim (all I+ & I–)	0.041	0.011	0.320
Fractional partial bias	–0.020	–0.007	–0.151
Total number of observations	182280	7175	20670
Total number unique	29512	1136	4114
Mean((I)/sd(I))	13.1	46.8	2.4
Completeness	91.9	98.9	89.7
Multiplicity	6.2	6.3	5.0

Table 3.6 ***SCALA scaling summary for TOSV N_6***

	Overall	InnerShell	OuterShell
Low resolution limit	52.83	52.83	2.90
High resolution limit	2.75	8.70	2.75
Rmerge	0.086	0.026	0.474
Rmerge in top intensity bin	0.028	–	–
Rmeas (within I+/I–)	0.121	0.037	0.670
Rmeas (all I+ & I–)	0.121	0.037	0.670
Rpim (within I+/I–)	0.086	0.026	0.474
Rpim (all I+ & I–)	0.086	0.026	0.474
Fractional partial bias	–0.044	–0.048	–0.278
Total number of observations	76447	2453	10787
Total number unique	39945	1267	5746
Mean((I)/sd(I))	6.8	22.2	1.7
Completeness	95.2	95.9	94.2
Multiplicity	1.9	1.9	1.9

Table 3.7 ***Crystallographic Summary***

Virus	**RVFV**	**RVFV**	**RVFV**	**RVFV**	**RVFV**	**TOSV**
Oligomeric State	**N_4-RNA_{28}**	**N_5-RNA_{35}**	**N_6-RNA_{35}**	**N_6-DNA_{30}**	**N_6**	**N_6**
Diffraction data						
Space group	*C*2	*P*1	*P*1	$P6_422$	$P3_121$	*P*1
Unit cell lengths (Å) a, b, c	77.6 193.2 77.4	79.8 93.6 124.7	91.6 173.3 172.9	108.6 108.6 261.3	107.1 107.1 258.4	50.3 93.7 95.5
Unit cell angles (°) α, β, γ	90.0 108.9 90.0	101.7 90.3 114.2	119.9 99.3 90.1	90.0 90.0 120.0	90.0 90.0 120.0	67.9 85.9 87.9
X-ray source	APS 23 ID-B	APS 23 ID-B	APS 23 ID-B	APS 23 ID-D	APS 23 ID-B	APS 23 ID-D
Wavelength (Å)	1.0332	1.0332	1.0332	1.0332	1.0332	1.0332
d_{min} (Å)	2.15 (2.27-2.15)[a]	3.90 (4.11-3.90)	3.40 (3.58-3.40)	2.70 (2.85-2.70)	3.10 (3.27-3.10)	2.75 (2.90-2.75)
Unique reflections	55,733	28,961	117,933	25,771	32,046	39,945
R_{merge}[b]	0.11 (0.58)	0.21 (0.46)	0.19 (0.51)	0.11 (0.78)	0.09 (0.69)	0.08 (0.47)
Avg I/σ_I	9.3 (2.3)	2.8 (1.6)	4.0 (1.7)	10.0 (2.0)	13.1 (2.4)	6.8 (1.7)
Completeness (%)	95.5 (77.9)	99.1 (99.1)	94.7 (92.3)	99.6 (98.5)	91.9 (89.7)	95.2 (94.2)
Average redundancy	3.7 (3.3)	2.0 (2.0)	1.9 (1.9)	4.4 (4.3)	6.2 (5.0)	1.9 (1.9)
Refinement						
Data range (Å)	31.48-2.15	41.6-3.90	57.42-3.40	53.66-2.70	46.4-3.10	50.22-2.75
Reflections	55,716	28,958	117,932	25,749	31,959	39,944
R/R_{free}[c]	0.180/0.219	0.228/0.248[d]	0.224/0.240[d]	0.189/0.230	0.215/0.256	0.218/0.238
RMS deviations						
Bond (Å)	0.009	0.008	0.008	0.010	0.009	0.009
Angles (°)	0.93	0.87	0.85	1.06	1.02	0.99
Avg B-factor (Å^2)						
Protein	39.6	56.8	70.5	60.2	122.5	83.0
Water	37.2	NA	NA	45.8	NA	28.4
DNA/RNA	36.0	36.8	117.2	81.6	NA	NA
Ramachandran[e]						
Allowed (%)	99.38	98.90	98.80	99.58	99.10	99.50
Outlier (%)	0.62	1.10	1.20	0.42	0.90	0.50
Number of Atoms						
Protein	7,636	19,093	67,752	5,698	11,404	11,221
Water	545	0	0	56	0	18
DNA/RNA	559	1400	4320	277	NA	NA
ASU Content	$(N_2\text{-}RNA_{14})_2$	$2x(N_5\text{-}RNA_{35})$	$6x(N_6\text{-}RNA_{36})$	$N_3\text{-}DNA_{13}$	N_6	N_6
PDB Code	XXX	XXX	XXX	XXX	XXX	XXX

[a] Values in parentheses are for the outermost shell of data.

[b] Rmerge = $\sum|Ii-\langle I\rangle|/\sum Ii$, where Ii is the intensity of the ith observation and <I> is the mean intensity. Sums are taken over all reflections.

[c] $R = \sum||F_o|-|F_c||/\sum|F_o|$. R_{free} is calculated for a 5% subset of the data

[d] For structures with 10 or more N subunits in the asymmetric unit, reflections in the R_{free} set were chosen in thin shells.

[e] Calculated with MolProbity (119).

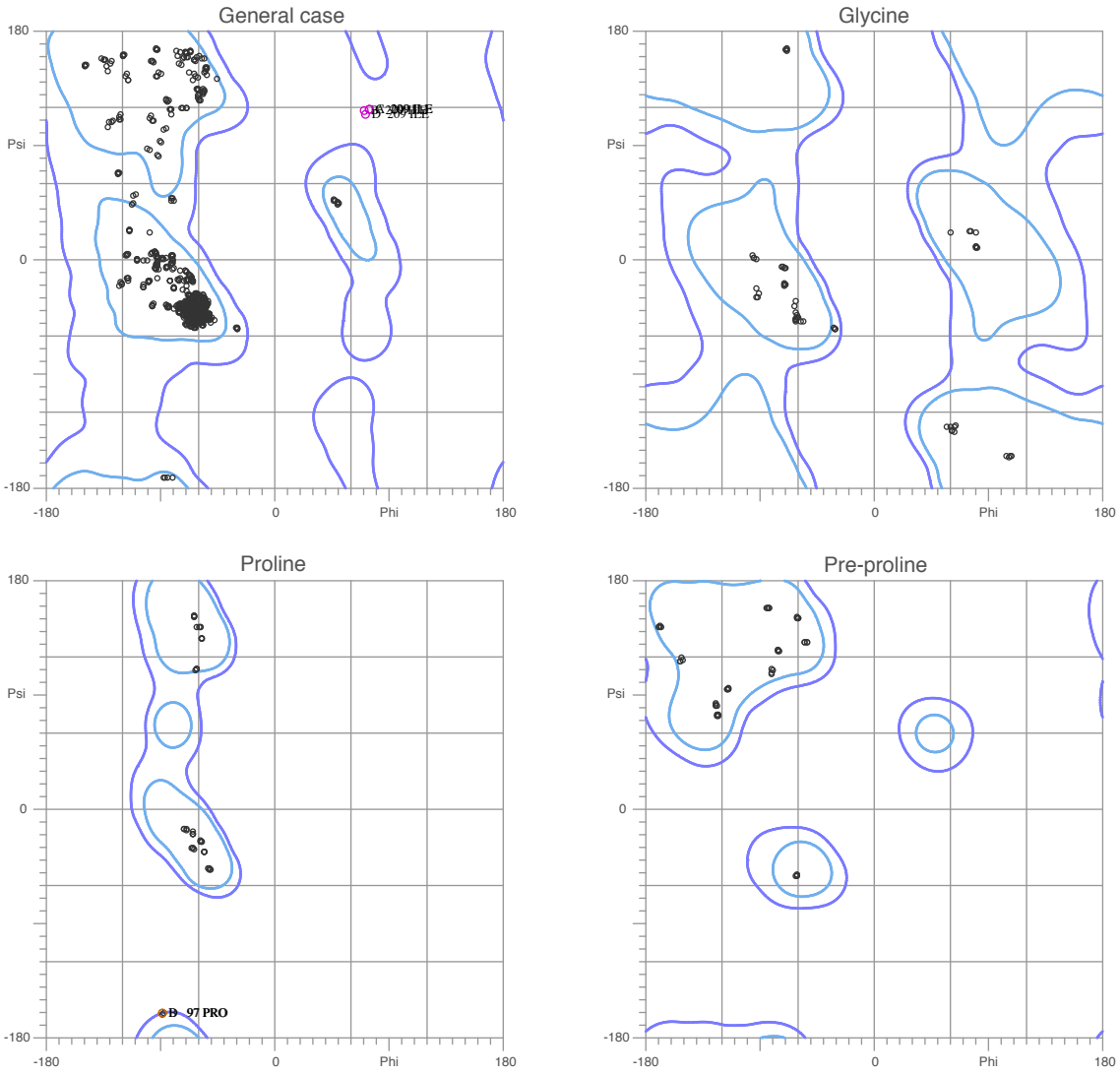


Figure 3.1 ***Ramachandran plot for RVFV N_4-RNA_{28}.***
As output from MolProbity, 99.4% of all residues were in the allowed regions.

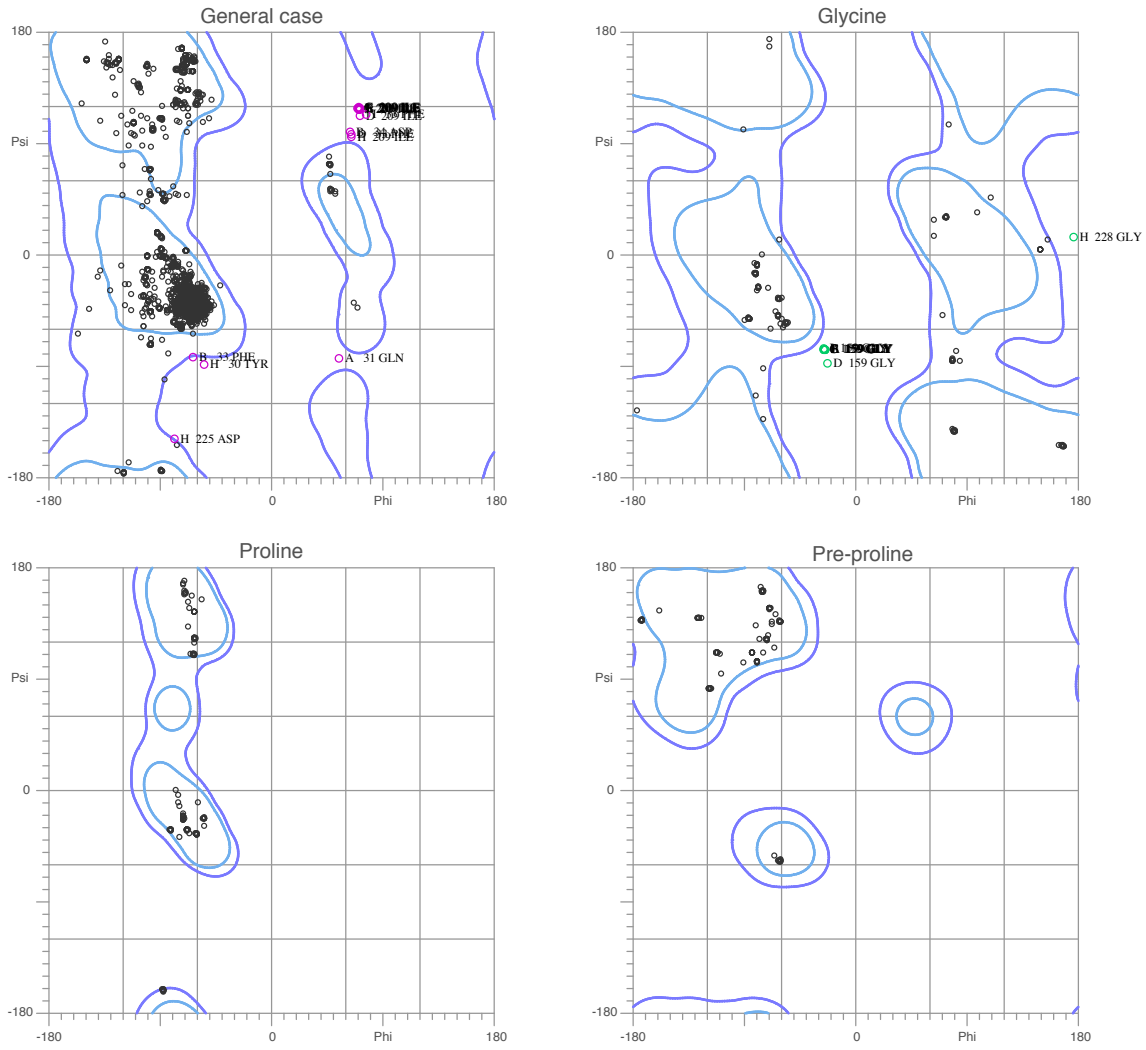


Figure 3.2 ***Ramachandran plot for RVFV N_5-RNA_{35}.***
As output from MolProbity, 98.6% of all residues were in the allowed regions.

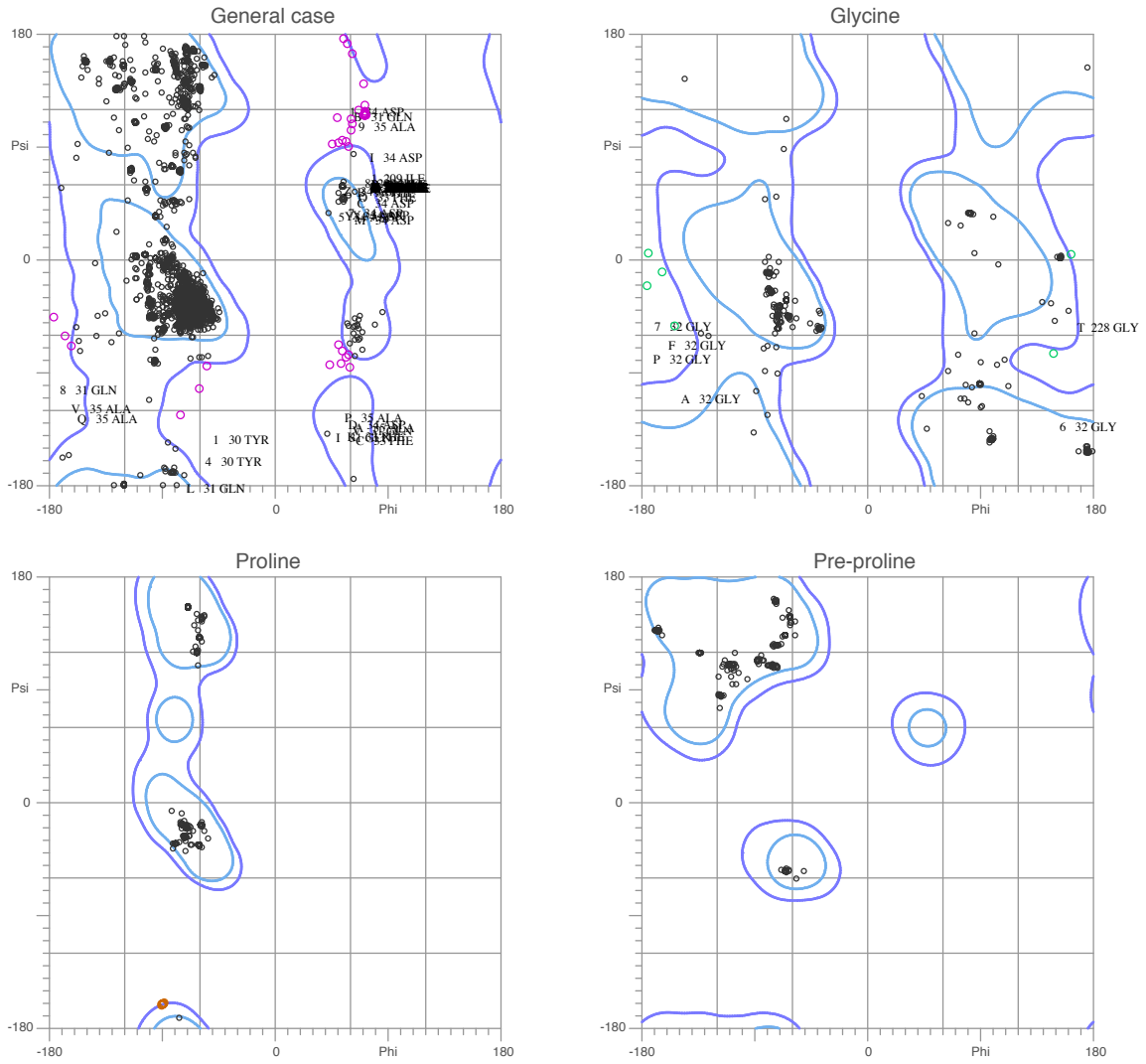


Figure 3.3 ***Ramachandran plot for RVFV N_6-RNA_{35}.***
As output from MolProbity, 98.8% of all residues were in the allowed regions.

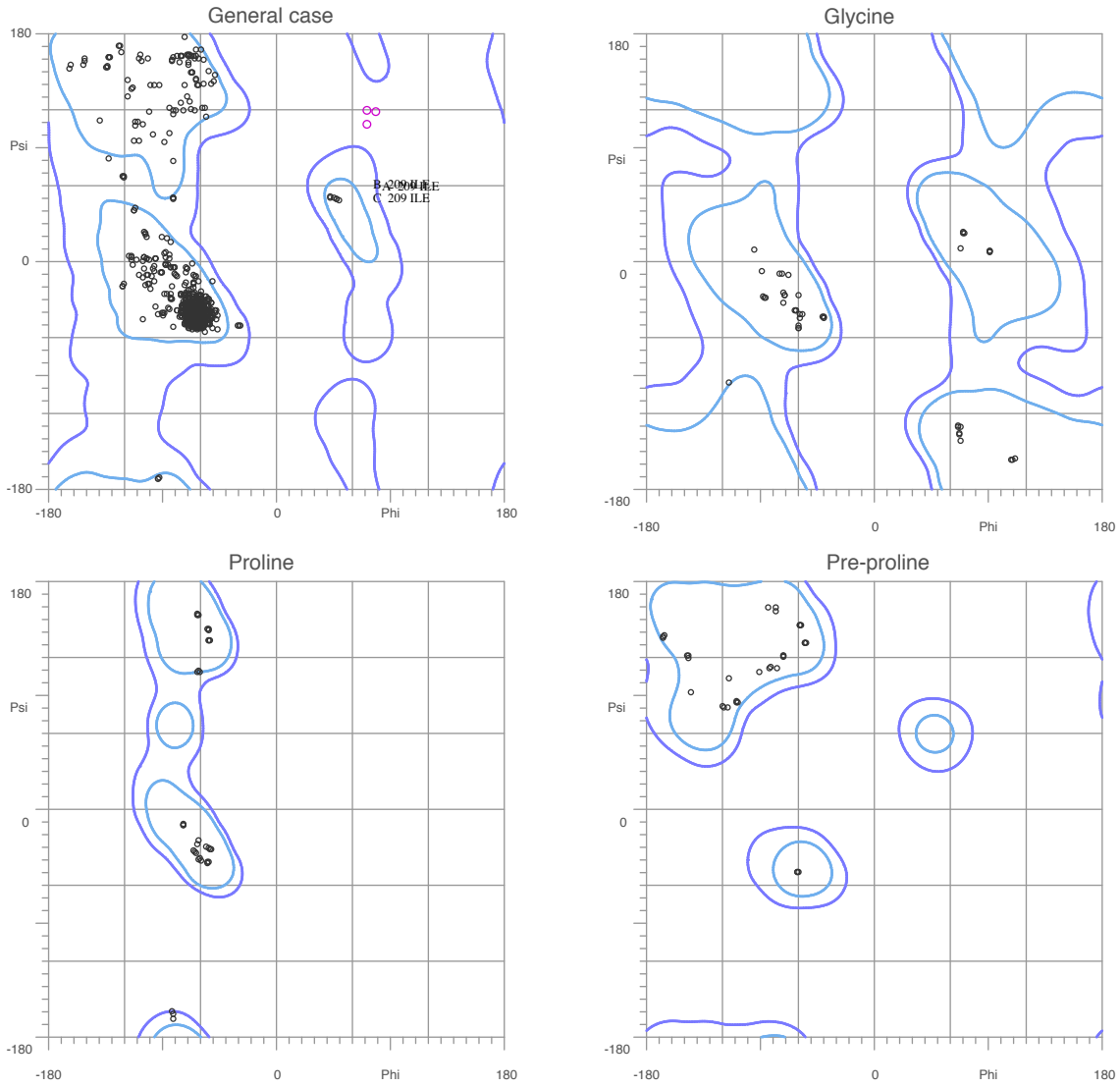


Figure 3.4 ***Ramachandran plot for RVFV N_6-DNA_{30}.***
As output from MolProbity, 99.1% of all residues were in the allowed regions.

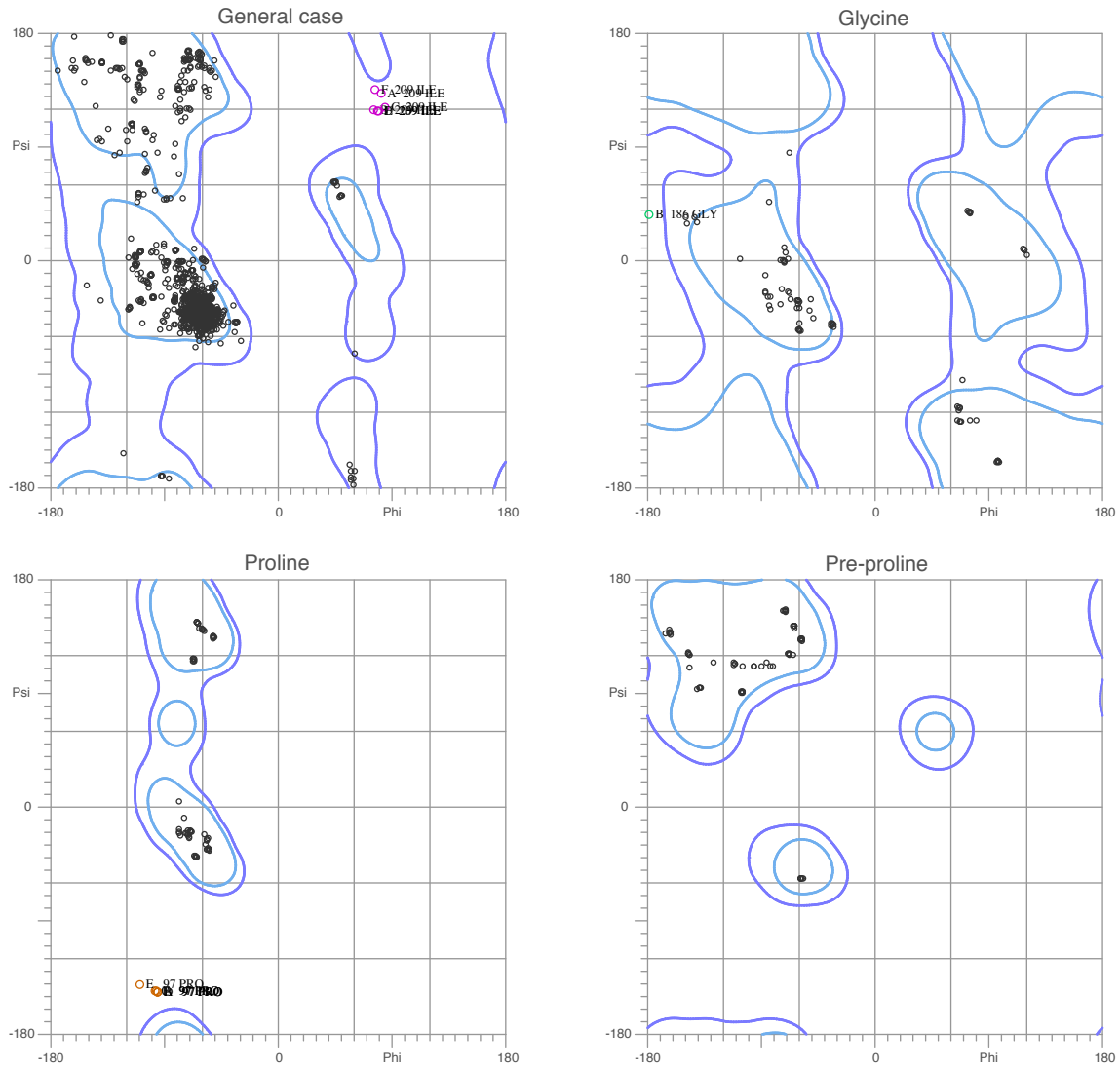


Figure 3.5 ***Ramachandran plot for RVFV N$_6$.***
As output from MolProbity, 99.1% of all residues were in the allowed regions.

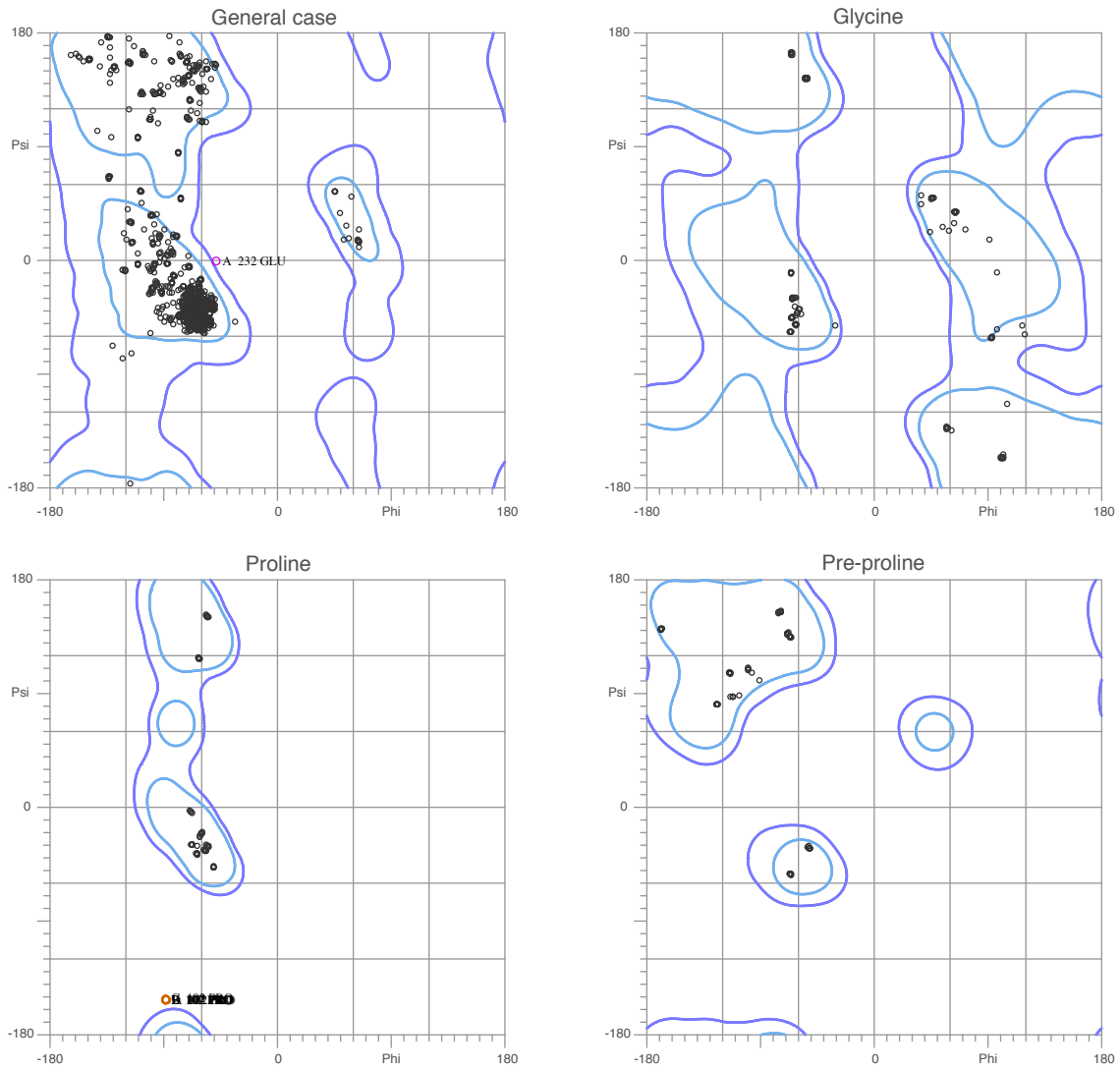


Figure 3.6 ***Ramachandran plot for TOSV N_6.***
As output from MolProbity, 99.5% of all residues were in the allowed regions.

Results

TOSV N behavior in solution

TOSV N behaves differently from RVFV N in solution. Gel-filtration analysis of natively purified and refolded recombinant TOSV N showed complexes of similar size to the multimeric RVFV N-RNA (Fig. 3.7). The chromatogram from Superdex 200 (S200) gel-filtration show a large 100 kDa peak and a broad shoulder to the left of the peak suggesting higher multimeric states of N (Fig 3.7). Fractions from natively-purified and RNA-free N had absorbance ratios of 1.25 and 0.56, respectively. In contrast to RVFV N purifications, TOSV N does not exist in mono- or dimeric states in solution (Fig 3.7).

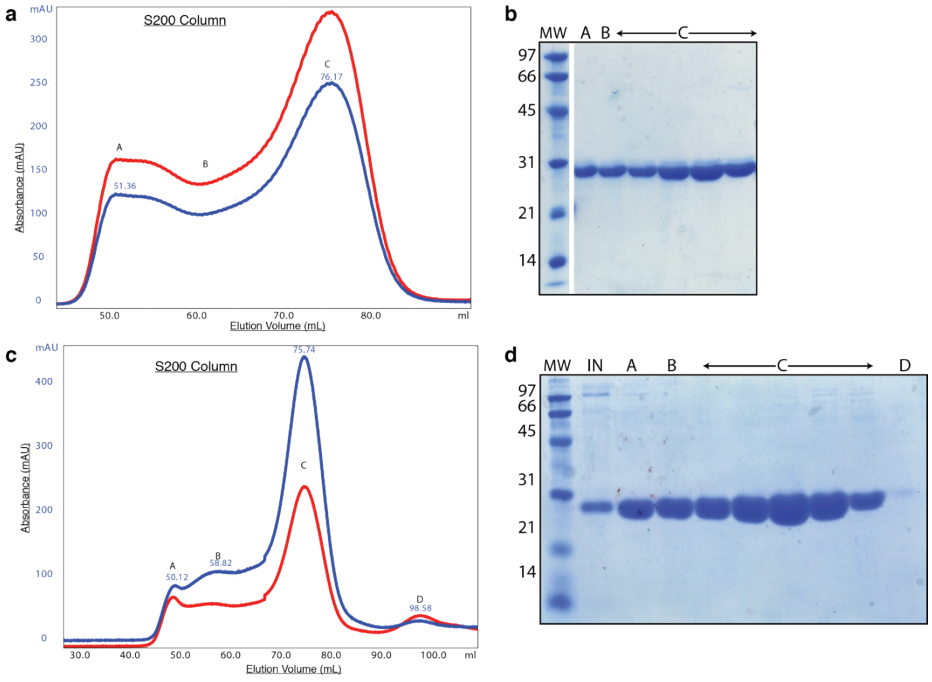


Figure 3.7 ***Purification of recombinant TOSV N.***
(**a**) Preparative S200 gel-filtration chromatogram of recombinant N after purification by Ni-affinity chromatography and cleavage of the SUMO fusion partner. Peak A is the void volume, peak B and C are N-RNA multimers. The red and blue traces represent absorption at 260 nm and 280 nm, respectively. (**b**) SDS-PAGE of fractions from the chromatogram shown in (**a**). Lane 1: molecular weight markers, lane 2-7: fractions from peak A-C, as labeled. (**c**) Preparative S75 gel-filtration chromatogram of refolded TOSV N. Peak A is in the void volume, peak B and C are N multimers, and peak D is free RNA. (**d**) SDS-PAGE of fractions from the chromatogram shown in (c). Lane 1: molecular weight markers, lane 2: input sample, lanes 3–10: fractions from peaks A–D, as labeled.

RVFV and TOSV N binds RNA and DNA nonspecifically

RNA-free N bound with high affinity to single-stranded nucleic acid, based on measurement of binding affinities by fluorescence polarization using labeled single-stranded oligomers of RNA or DNA (Table 3.8 & Fig. 3.8). The RNA-free N from two phleboviruses, RVFV and TOSV had similar affinities (6-26 nM) for RNA and DNA oligomers over a broad size range (10-35 nucleotides). RVFV and TOSV RNA-free N also had similar binding affinities for RNA and DNA strands with mixed sequences (e.g. $(ACC)_8A$) (Fig 3.8a & b). In this direct test of affinity for nucleic acid, RNA-free N displayed no sequence specificity. The high affinity for DNA is likely irrelevant to biological function, as N is not thought to enter the nucleus (30). Phleboviruses replicate in the host cytoplasm and thus have no need to discriminate RNA from DNA. The only experimental factor that affected N affinity for RNA was the nucleic acid content of the N preparation, as judged by the ratio of absorbances at 260 nm and 280 nm. (Table 3.8 & Fig. 3.8c).

Table 3.8 ***RVFV N Binding affinities for single-stranded nucleic acid.***

Type	Name	Sequence	Length	Protein	A_{260}/A_{280}	k_d (nM)
RNA	U_{25}	U_{25}	25	RVFV Recombinant N monomer*	0.75	96 ± 14
RNA	U_{25}	U_{25}	25	RVFV Recombinant N multimer*	1.32	> 4,800
RNA	U_{25}	U_{25}	25	RVFV RNA-free N	0.56	13 ± 1
RNA	U_{30}	U_{30}	30	RVFV RNA-free N	0.56	9 ± 2
RNA	U_{35}	U_{35}	35	RVFV RNA-free N	0.56	11 ± 2
RNA	RNA_{10}	AAC GCU UCC C	10	RVFV RNA-free N	0.56	10 ± 3
DNA	T_{10}	dT_{10}	10	RVFV RNA-free N	0.56	13 ± 3
DNA	T_{25}	dT_{25}	25	RVFV RNA-free N	0.56	14 ± 3
DNA	T_{35}	dT_{35}	35	RVFV RNA-free N	0.56	26 ± 3
DNA	DNA_{25}	$d(ACC)_8$ A	25	RVFV RNA-free N	0.56	18 ± 3
DNA	T_{25}	dT_{25}	25	TOSV RNA-free N	0.56	6 ± 2
RNA	U_{25}	U_{25}	25	TOSV RNA-free N	0.56	14 ± 3

RNA-free N was generated by denaturation and refolding.
*Recombinant RVFV N yielded both monomeric and multimeric species after digestion with ribonuclease.

Reconstitution and crystallization of N-nucleic acid complexes

The equal-affinity binding of N to nucleic acid oligomers of different lengths is consistent with the appearance of N-RNA or N-DNA in negative stain EM visualization experiments (145, 146) (Fig. 3.9a & b). The reconstituted N-nucleic acid complexes appeared identical to both authentic RNP from virus-infected cells following extensive ribonuclease treatment and recombinant N-RNA that had not been stripped of RNA (145). The mixed population of multimers appeared to form

predominantly closed ring structures, but some appeared as incomplete or open rings (Figs 2.6 & 3.9c). Therefore, to select a nucleic acid target for crystallization, we used negative stain EM to screen oligomers for the length that generated the most homogeneous multimer population. Previous experiments showed that N pentamers and hexamers were the most prevalent species of digested authentic viral RNP and of recombinant N-RNA (145, 146). Additionally, the electrophoretic mobility of RNA extracted from recombinant N-RNA had a broad size distribution centered between 30 and 40 nucleotides (Fig. 3.10), consistent with the length of a positively charged cleft on the inner surface of an observed hexameric N (146). We thus screened complexes that were reconstituted with DNA oligomers of length 25-45 nucleotides. The N-DNA_{35} complex yielded the most homogeneous multimer population, with ring-like structures of predominantly five or six N subunits (Fig. 3.9c & d).

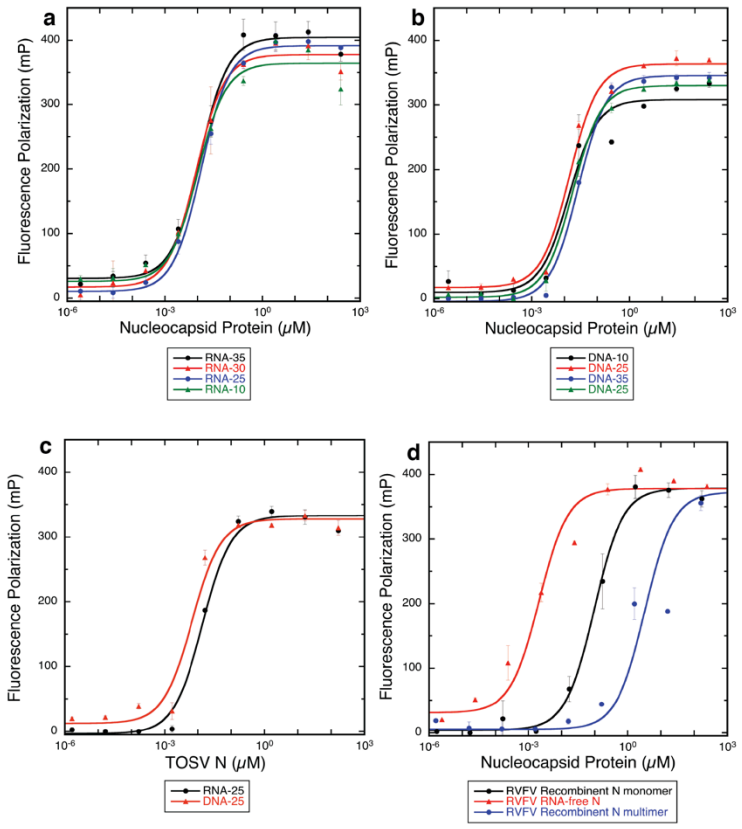


Figure 3.8 ***Fluorescence polarization measurement of N binding to nucleic acid.***
(**a**) RNA-free N binding to RNA. (**b**) RNA-free N binding to DNA. (**c**) RNA-free TOSV N binding to RNA and DNA. (**d**) RNA_{25} binding to N preparations with varying RNA content.

Reconstituted RVFV N-DNA and N-RNA complexes were purified by gel filtration and subjected to crystallization screening, along with RNA-free N from RVFV and TOSV. Crystals were obtained for RVFV N-RNA, RVFV N-DNA, RVFV RNA-free N, and TOSV N (49% identical to RVFV N). Initial N-DNA_{35} crystals diffracted to 3.6 Å, but further optimization of the crystallization conditions and DNA length and purity resulted in N-DNA_{30} crystals with a diffraction limit of 2.7 Å (Table 3.8). Two crystal forms of N-RNA_{35} and one of N-RNA_{28} diffracted to 3.9 Å, 3.4 Å and 2.15 Å, respectively.

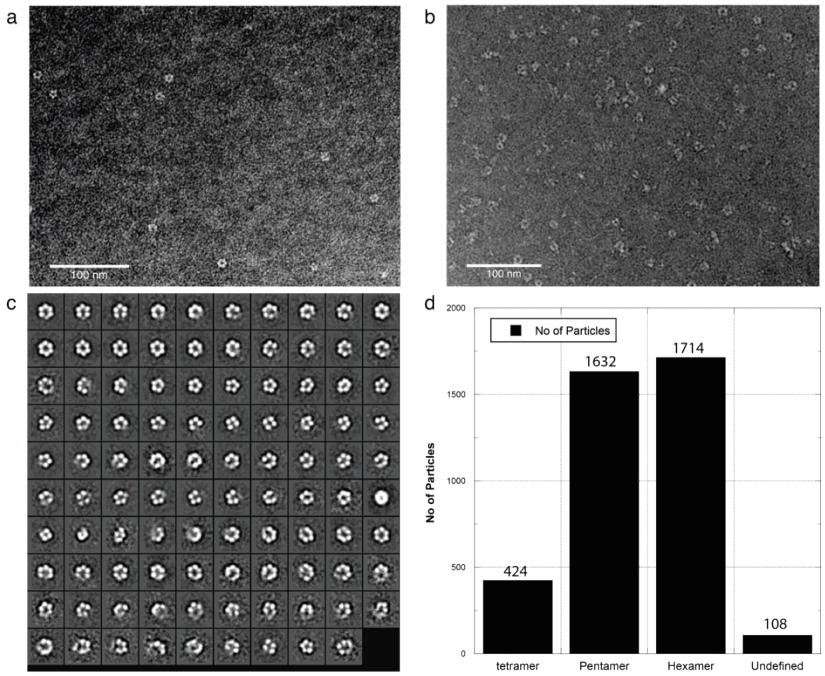


Figure 3.9 ***Electron microscopy of RVFV N-DNA*****.**
Comparative EM fields of negatively stained (a) N-DNA_{35} and (b) N-DNA_{40} complexes. (c) Reference-free classifications of 3770 N-DNA_{35} particles in 99 classes. (d) Histogram of oligomeric state distribution of N-DNA_{35} particles.

RVFV and TOSV N Model quality and electron density

The final models of RVFV and TOSV N multimeric structures were determined at resolutions from 3.9 to 2.15 Å with reasonable R/R_{free} values (Table 3.8). MolProbity analysis of the refined models showed good geometric quality and low clash scores (>96^{th} percentile of structures in each resolution range) (119). Ramachandran plots by MolProbilty showed >98% of residues in allowed regions for all structures (Figs

3.1-6). Continuous electron density was observed for residues 2-245 in all RVFV multimers and residues 3-253 in the six subunits of the TOSV N hexamer. A three-residue insertion at the end of α2 in TOSV N was visible in the electron density. In general, the refined models had adequate electron density for model building, and the high-resolution structure of the RVFV N monomer facilitated accurate placement of amino-acid residues during the building process.

N multimer formation

All crystals, with or without nucleic acid, contained N multimers (Figs 3.11-3.13). RVFV RNA-free N, N-DNA_{30} and TOSV N crystallized as hexamers (N_6 or N_6-DNA_{30}). N-RNA_{35} crystallized in identical conditions as either a hexamer (N_6-RNA_{35}) or a pentamer (N_5-RNA_{35}), whereas N-RNA_{28} crystallized as a tetramer (N_4-RNA_{28}). The 65 independent views of the N subunit from the six structures (Table 3.8) and seven views from two published structures (145, 146) are identical in the subunit core, but they differ in the position of the helical arm (α1 and α2), which extends from the core of each subunit to contact the core of a neighbor (146) (Figs 3.11-13). The arm-to-core interactions are essentially identical in all subunits of all crystal structures, and account for nearly all protein-protein contacts in the N multimer (Figs 3.11-13). Remarkably, the crystallized multimers deviate from perfect four-, five- or six-fold symmetry.

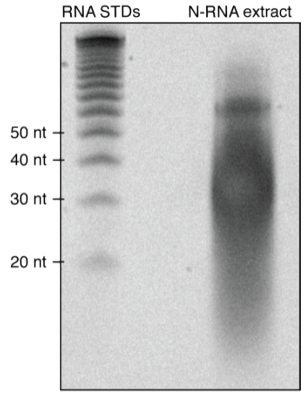


Figure 3.10 ***Analysis of RNA extracted from recombinant RVFV N multimer*****.**
After extensive ribonuclease digestion, RNA was extracted from purified recombinant N multimer with an equal-volume mixture of phenol and chloroform. The extracted RNA was separated on a denaturing urea-polyacrylamide gel and visualized using SYBR Green. Single-stranded RNA standards are shown in the first lane. The RNA was extracted and analyzed by Mary Piper.

Nucleic acid electron density

In all N-nucleic acid structures, the multimer unit is a ring of N subunits. The nucleic acids used in crystallization were homo-oligomers (poly-U or poly-dT) that lacked terminal phosphates. Positive difference density corresponding to nucleic acid was observed in structures of reconstituted N-nucleic acid in maps calculated with phases from molecular replacement with an RVFV N probe (Fig. 3.14). The ends of the nucleic acid oligomers distributed randomly within the multimers, so electron density was continuous around the RNA-binding slot with no visible termini (Fig. 3.15). In the N_4-RNA_{28} and N_5-RNA_{35} structures, random binding would result in an undetectable reduced occupancy at the phosphate positions by 1/28 and 1/35, respectively, as each N subunit bound 7 nucleotides. In the N_6-RNA_{35} structure, nucleotides were observed in 36 positions because each N subunit bound 6 nucleotides. Although TOSV N was purified under native conditions and the A_{260}/A_{280} ratio of 1.25 indicated RNA was present in the N sample, no density corresponding to RNA was observed in the crystal structure (Figs 3.13).

N RNA binding

RVFV N binds single-stranded RNA using an unusual base sequestration mechanism. Each N subunit possesses a deep, narrow RNA-binding slot, which becomes a continuous groove on the inner surface of the ring of 4-6 N subunits (Figs 3.16-3.17). The inner surface of the groove is lined with conserved hydrophobic amino acids while the rim has several conserved positively charged residues (Figs 3.17-3.18). Nucleic acid binds with the bases inserted into the slot and the sugar-phosphate backbone oriented towards the solvent in the center of the N multimer (Figs 3.11-3.13 & 3.17-3.18). The high affinity of N for single-stranded nucleic acid is explained by extensive hydrophobic contacts of bases with amino acids in the RNA-binding slot and by base stacking. Although we crystallized N with oligomers of pyrimidine nucleotides, the RNA-binding slot is deep enough to accommodate purines (Fig. 3.17).

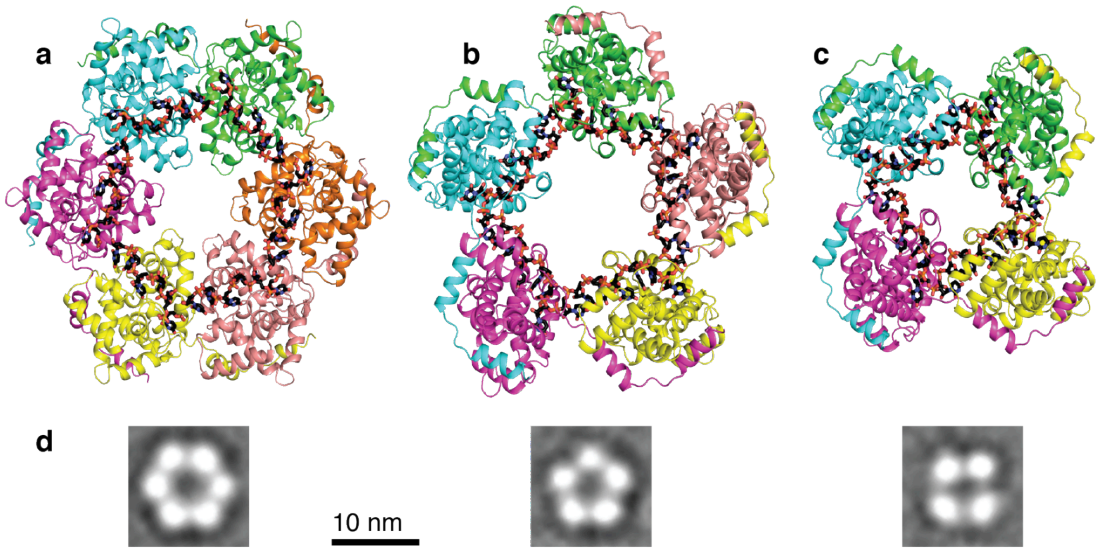


Figure 3.11 ***Structural plasticity of phlebovirus N-RNA.***
(**a**) N_6-RNA_{35}, (**b**) N_5-RNA_{35}, and (**c**) N_4-RNA_{28}. N subunits are rendered in ribbon form with contrasting colors. The helical arm of each subunit wraps around the neighboring subunit on the outside of the multimer. Single-stranded RNA, rendered in stick form with black carbon atoms, binds to the inner surface of the multimer with all bases pointing into the protein. (**d**) EM visualization of reconstituted hexamer, pentamer and tetramer multimers. The class averages, from Fig. 3.9c, are representative of the multimers released from viral RNP by ribonuclease digestion(145).

Each N subunit binds four RNA nucleotides in the RNA-binding slot and 2-3 additional nucleotides in the subunit interface. RNA is secured into the binding slot by contacts with 18 conserved amino acids (Figs 3.14 & 3.18). In each N subunit, the 5′-most base (B1) stacks with Tyr30 in the hinge region between the helical arm and the N core (Figs 3.14 & 3.18). B2 stacks with Phe33 in a "back pocket" of the RNA-binding slot (Figs 3.17-3.18). The B3 and B4 bases are stacked in the central compartment of the RNA-binding slot, which is lined with the side chains of Ala109, Ala110, Pro147, Ile180, Pro199 and Ala202. B5 occupies a narrow pocket lined with the side chains of Gly 65, Leu126, Pro127 and Phe176 (Fig. 3.18). B6 and B7 stack with B1′ at the downstream N-N′ interface in the N_5-RNA_{35} pentamer and in the N_4-RNA_{28} tetramer (Fig. 3.19a). The N_6-RNA_{35} hexamer lacks a B7, and B6 is stacked with B1′ (Fig. 3.19b). At the N-N interface, the positions of B1 and B2 create a sharp bend in the RNA backbone, which is more pronounced in the tetramer structure (~80°) than in the pentamer (~100°) and hexamer (~105°) structures (Fig. 3.11). In addition to the hydrophobic and base-stacking interactions, N forms a network of

polar contacts with the RNA 5′ phosphates (P1-P7) of all nucleotides except P1: the Phe33 amide and Tyr30 hydroxyl group with P2; Arg99 with P2 and P3; Arg106 with P3; Asn66 with P4 and the 2′-hydroxyl group of R5, Lys67 with P5 and P6; and Arg70 with P6 and P7 (Figs 3.14 & 3.18). Consistent with sequence-independent RNA binding, the N subunit forms no H-bonds to RNA bases.

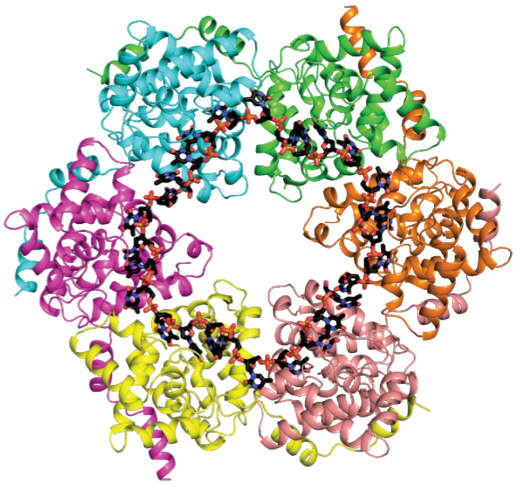


Figure 3.12 ***Crystal structure of RVFV N6-DNA30*.**
N subunits are rendered in ribbon form with contrasting colors. The helical arm of each subunit wraps around the neighboring subunit on the outside of the multimer.

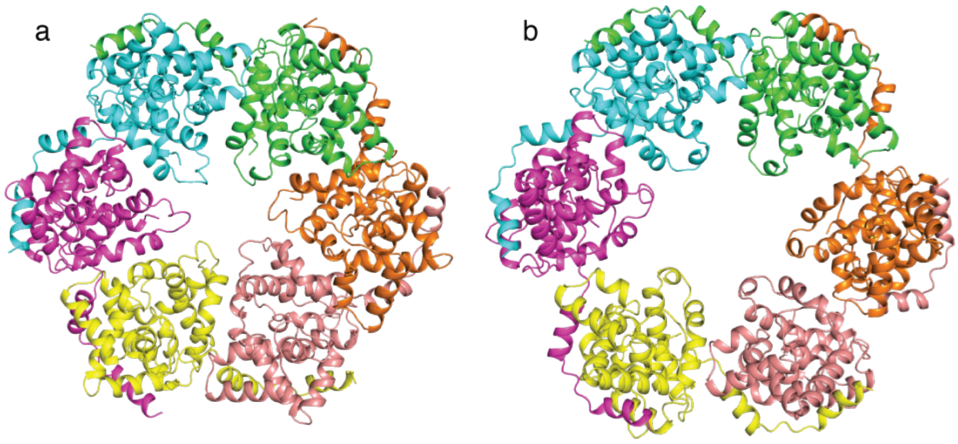


Figure 3.13 ***Structures of phlebovirus N*.**
(**a**) RVFV N_6 and (**b**) TOSV N_6. N subunits are rendered in ribbon form with contrasting colors. The helical arm of each subunit wraps around the neighboring subunit on the outside of the multimer. Note that neither the RVFV nor the TOSV N hexamer is perfectly six-fold symmetric.

DNA binding in the RNA-binding slot

DNA and RNA bound somewhat differently in the RNA-binding slot. The number of DNA nucleotides bound to each N subunit varied. Density, although continuous, was visible for only 28 of 30 nucleotides (Fig. 3.12). Four subunits of the N_6-DNA_{30} hexamer had no nucleotide in position 2 (Fig. 3.18). In the other two subunits, B2 occupied a different position within the RNA-binding slot and did not stack with Phe33 (Fig. 3.18). At the interface of two N subunits; the N_6-DNA_{30} hexamer lacked nucleotides at both positions 6 and 7 whereas, in the N-RNA hexamer and pentamer structures, B6 and B7 stack with B1′ (Fig. 3.19). There were also significantly fewer interactions with the DNA sugar-phosphate backbone in all subunits of the N_6-DNA_{30} hexamer. Only three of the eight backbone interactions seen in the N-RNA structures were observed in the N_6-DNA_{30} hexamer: Lys67 and the Asn66 main-chain amide to P5 and Lys70 to P1′ (Fig. 3.18). Despite these differences, N had equal affinities for RNA and DNA (Table 3.7 & Fig. 3.18). Density was visible for only 28 nucleotides in N_6-DNA_{30}. Because the hexamer sits on a crystallographic 2-fold, there are two gaps in electron density corresponding to the chain ends.

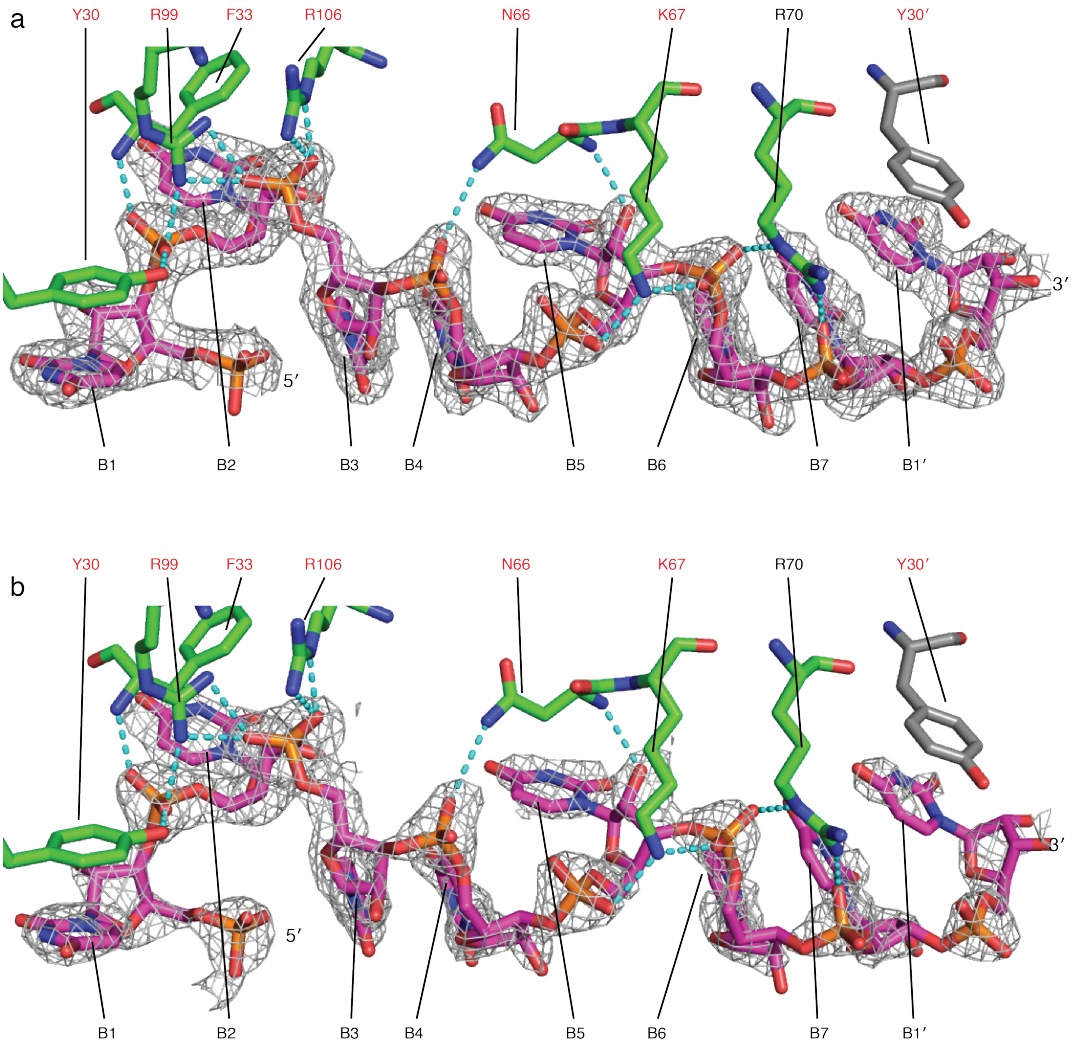


Figure 3.14 ***RVFV N RNA-binding interactions in the N_4-RNA_{28} tetramer.***
Amino acids and RNA nucleotides are rendered in stick form. The carbon atoms in amino acids from adjacent subunits are colored green and gray. The 2.15 Å (a) $2F_o$-F_c map contoured at 1σ and (b) F_o-F_c omit map contoured at 3σ of RNA bound to an N subunit. Amino acids labeled in red are invariant in phlebovirus N proteins. RNA bases 1-7 are labeled B1-B7 and B1′ is B1 of the adjacent N subunit.

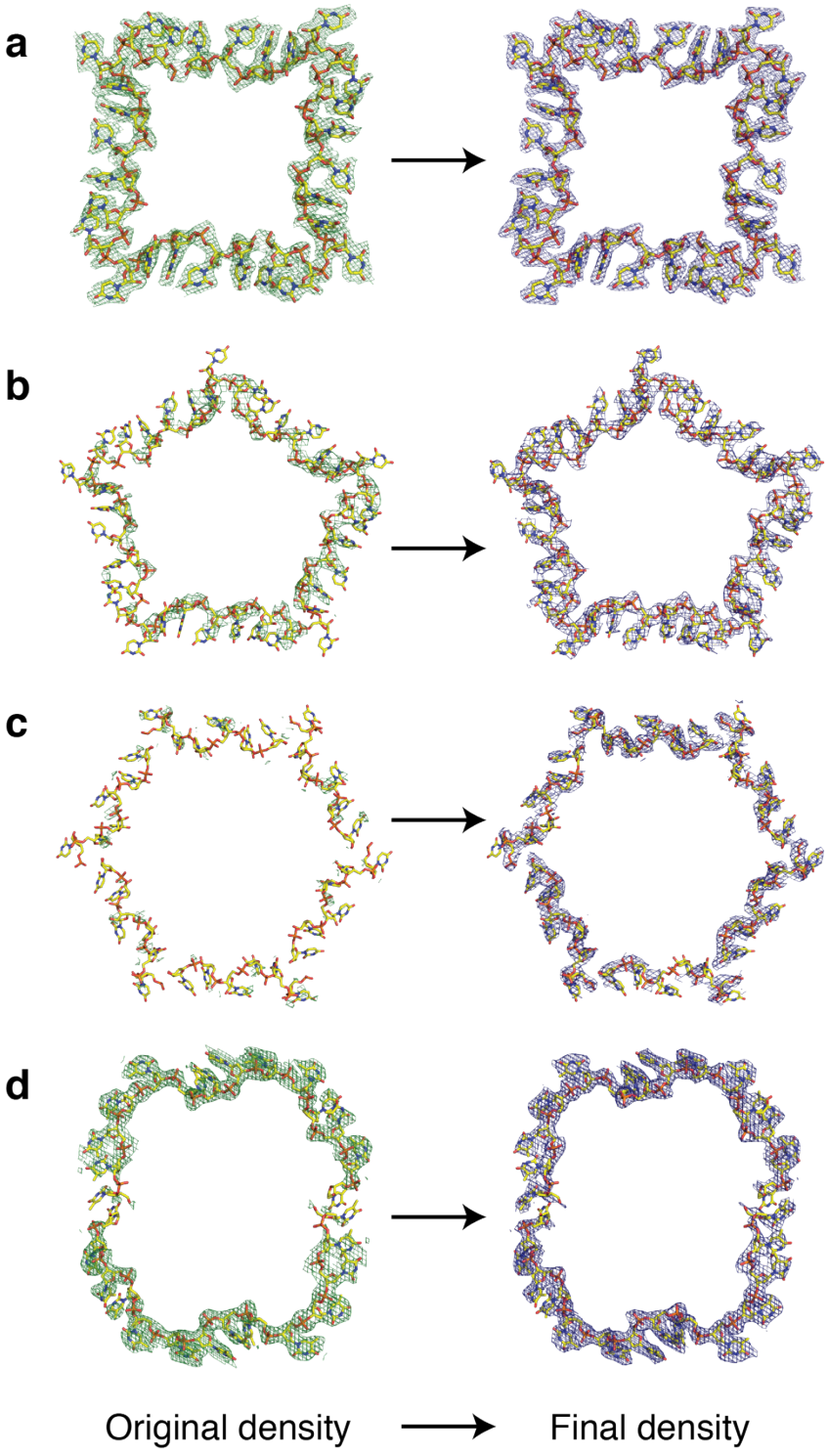


Figure 3.15 ***Electron density of nucleic acid bound to N.***
Electron density before (F_o-F_c contoured at 2.5σ) and after ($2F_o$-F_c contoured at 1σ) refinement with nucleic acid in the RNA-binding slot in crystal structures of (**a**) N_4-RNA_{28} at 2.15 Å, (**b**) N_5-RNA_{35} at 3.90 Å, (**c**) N_6-RNA_{35} at 3.40 Å and (**d**) N_6-DNA_{30} at 2.70 Å. Nucleic acids in sticks form with yellow C atoms.

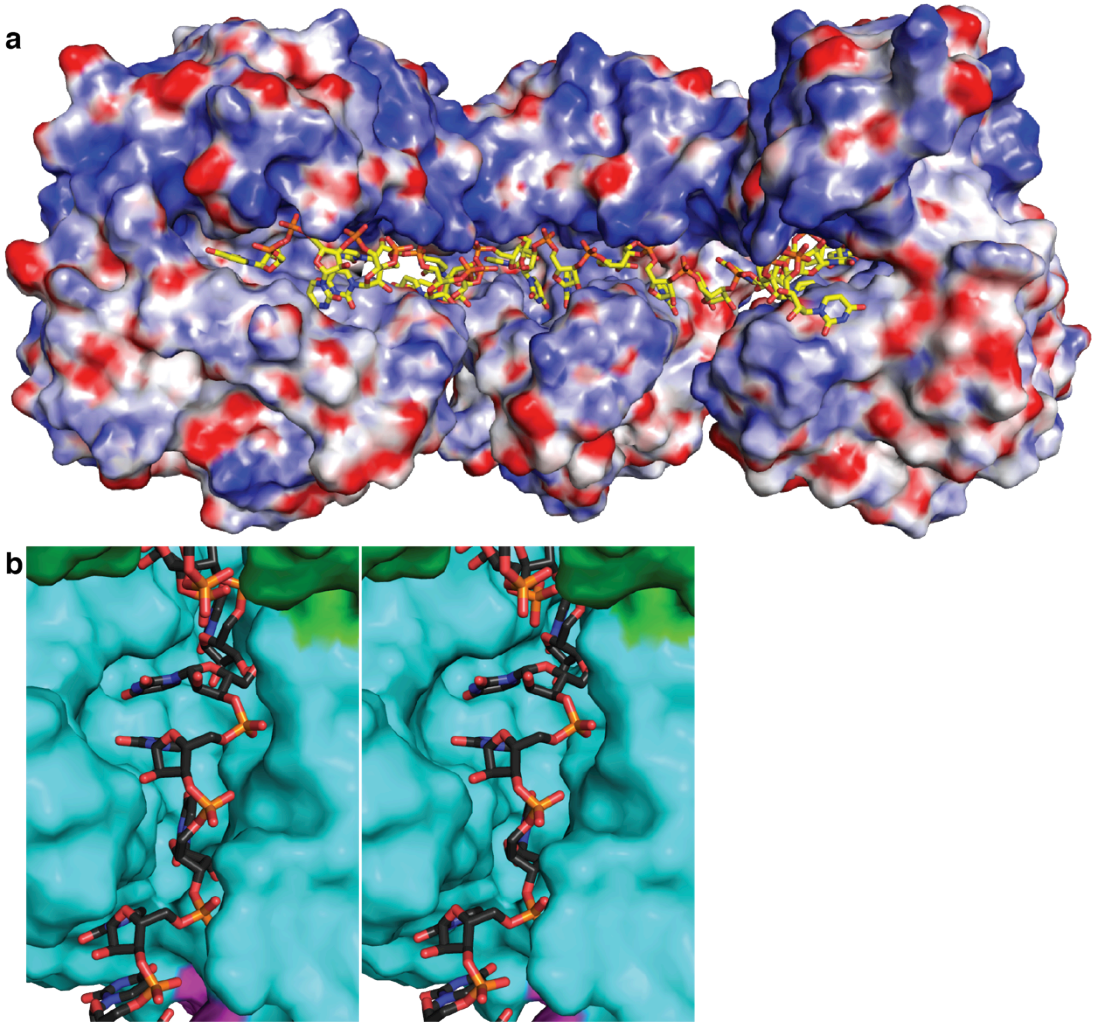


Figure 3.16 ***RNA sequestered in the RNA-binding slot***.
(**a**) The surface potential from -14 kT in red to +14 kT in blue is shown for three subunits in the N_5-RNA_{35} pentamer. The RNA is rendered in sticks form with yellow carbon atoms. (**b**) Stereo image of nucleotide binding in the RNA-binding slot of one N subunit. RNA is drawn in stick form with black C atoms and the surfaces of adjacent N subunits are colored green, cyan and magenta.

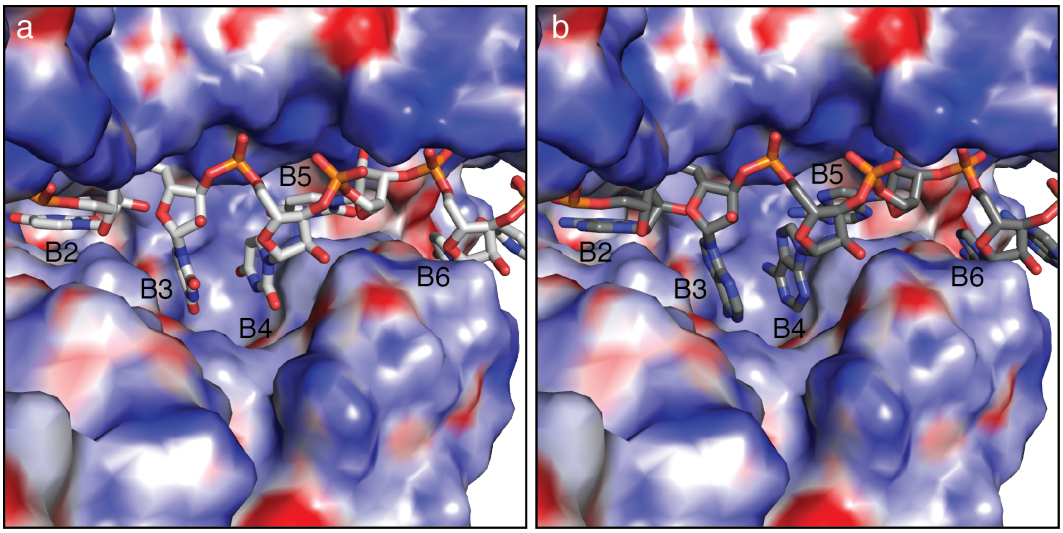


Figure 3.17 ***Pyrimidine and purine nucleotides in the RVFV N RNA-binding slot.***
(**a**) Observed uracil nucleotides from RVFV N_4-RNA_{28} and (**b**) modeled adenine nucleotides are rendered in stick form. The surface potential from -14 kT in red to +14 kT in blue is shown for the RNA-binding slot.

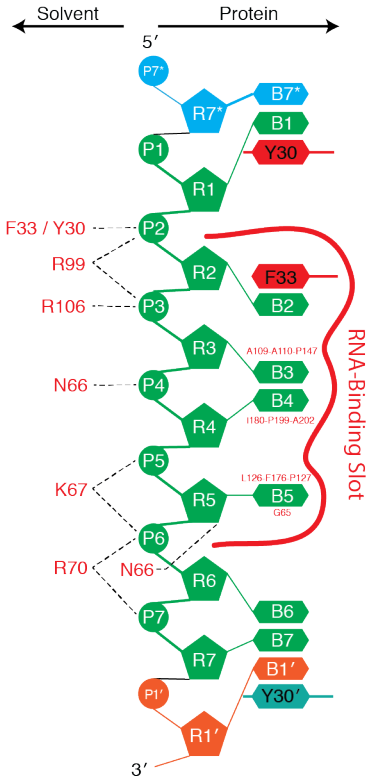


Figure 3.18 ***Schematic diagram of RVFV N interactions with RNA.***
P, R and B represent RNA phosphates, riboses and bases, respectively. Amino acids from adjacent N subunits are shown in red and cyan. RNA nucleotides in blue with a * label and in orange with ′ label interact with adjacent N subunits.

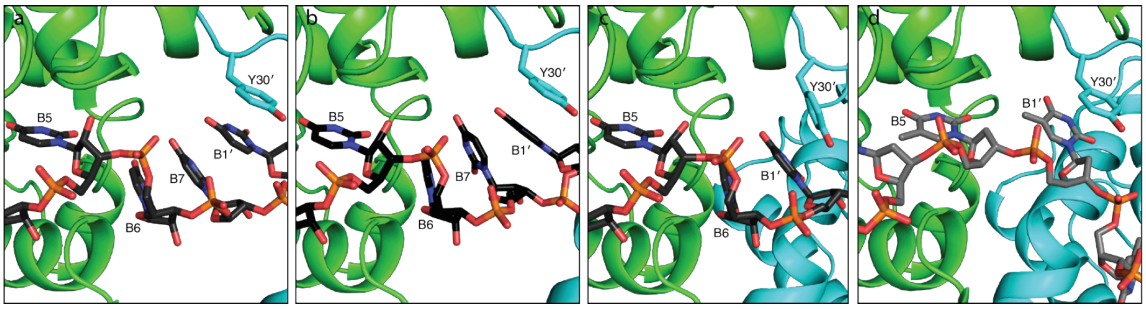


Figure 3.19 ***Base stacking with Tyr30 at the interface of RVFV N subunits.***
(**a**) N_4-RNA_{28}, (**b**) N_5-RNA_{35}, (**c**) N_6-RNA_{35} and (**d**) N_6-DNA30. Tyr30 and nucleic acid are rendered in stick form. N is rendered in ribbon form with adjacent subunits in green and cyan. The green subunit is viewed in the same orientation in all panels to illustrate the variability in orientation of the neighboring cyan subunit

Discussion

The crystal structures of reconstituted RVFV N-RNA complexes offer the first high-resolution view of genome packaging in segmented negative-sense RNA viruses. The unusual RNA-binding mechanism and limited N-N interactions in the RNP represent a new paradigm for RNP assembly. This work established three important properties of phlebovirus N: the mechanism of multimer formation for N proteins, non-specific sequestration of nucleotide bases in the RNA-binding slot, and variable base stacking between N subunits.

The ability to bind RNA is an essential function of N. N binds RNA non-specifically to ensure the complete encapsidation of the viral genome. Negative-sense viruses achieve sequence-independent RNA binding in at least three different ways: Base-in, base-out and mixed orientation. This study revealed that phleboviruses employ a base-in binding mode with a hydrophobic slot that is deep and long enough to sequester four pyrimidine or purine bases in each N subunit (Figs 3.16-3.18). The bases of nucleotides between N subunits are also directed towards the protein through stacking with Tyr30. None of the bases are available for "reading" by proteins or nucleic acids. In contrast to RVFV N, the Lassa virus N protein binds eight RNA nucleotides in a base-out orientation with the sugar-phosphate backbone directed into an RNA-binding pocket (153). RSV, VSV and rabies virus have a mixed binding mode where three stacked bases point into an RNA-binding pocket and another four or six bases face the solvent (93-95). In these viruses with partial or complete base-out RNA binding, the RdRp may be able to access the genome without disassembling the helical RNP (94, 95).

Strikingly, in RVFV, a four-nucleotide RNA core binds all N subunits identically regardless of multimer size. The RNA core includes nucleotides 2-5 and the flanking phosphates (Figs 3.14, 3.17, 3.18 & 3.20). Similarly, contacts of the helical arm of N with the neighboring subunit are identical in all subunits of all multimers (Fig. 3.21). These invariant RNA-protein and protein-protein contacts, which are located on opposite faces of the N core domain (Fig. 3.20), define a compact unit of identical structure, the "RNA-N_{core}-arm" unit, in all RNA-N multimers.

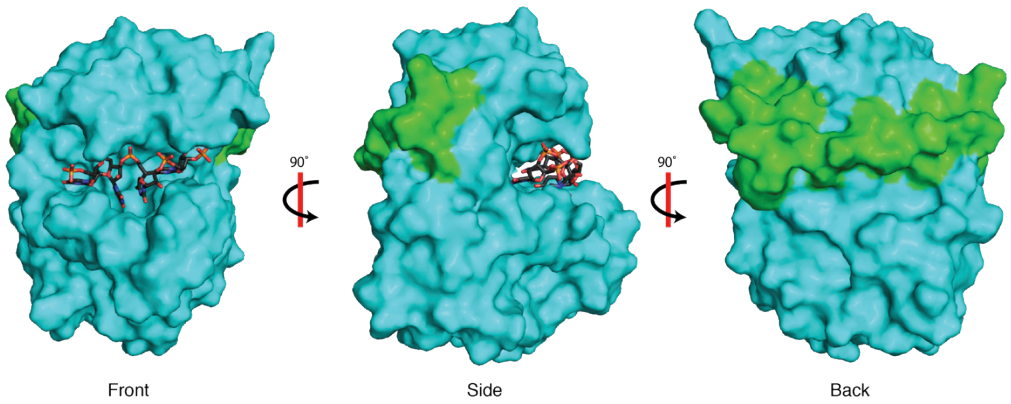


Figure 3.20 ***The RNA-N_{core}-arm building block of phlebovirus RNP.***
The core of an N subunit (cyan) binds the helical arm (green) of an adjacent N on one surface and four RNA nucleotides (sticks) on the opposite surface.

Despite the identical RNA-protein and protein-protein contacts, the structures vary considerably among and within multimers (Fig. 3.11). The variability derives from differences in the number of stacked bases between N subunits, in the hinge between the N core and its helical arm, and in the sharp bend between nucleotides 1 and 2. These structural features are located at the N-N subunit interface where a lack of contacts between N cores also facilitates variability (Fig. 3.19). The arm hinge (amino acids 28-35, including Tyr30 stacked on RNA B1) varies extensively among the multimers (Fig. 3.22). Hinge flexibility was evident in the earlier structure of an N hexamer (146), but the six new structures reveal a far greater range of arm motion (Figs 3.11-3.13).

The crystal structures of multimers show the RNA-protein interaction in exquisite detail, but all multimers are too large to be a building block for RNP (99, 145, 154) (Fig. 3.23). The RNP is a flexible chain of monomer-sized building blocks, consistent with the size of the invariant RNA-N_{core}-arm unit in all three N-RNA crystal structures (Fig. 3.23 inset). The irregular, asymmetric structure of phlebovirus RNP is fully explained by flexibly linked RNA-N_{core}-arm building blocks that can move relative to one another over the range captured in the crystal structures. The deep RNA-binding slot in each subunit encapsidates the RNA fully, including the nucleotides between subunits (Fig. 3.16). This rather primitive system for genome encapsidation requires few protein-protein contacts and sequesters all RNA bases

away from solvent and from access to other proteins. It is incompatible with base pairing, even for the nucleotides bound between N subunits. Furthermore, the structures explain the insensitivity of RNP to ribonuclease and to high salt treatment. The viral genome is protected by N in the ribonuclease-resistant RNP(145), which serves as the substrate for transcription and replication by the RNA-dependent RNA polymerase (RdRp) (31).

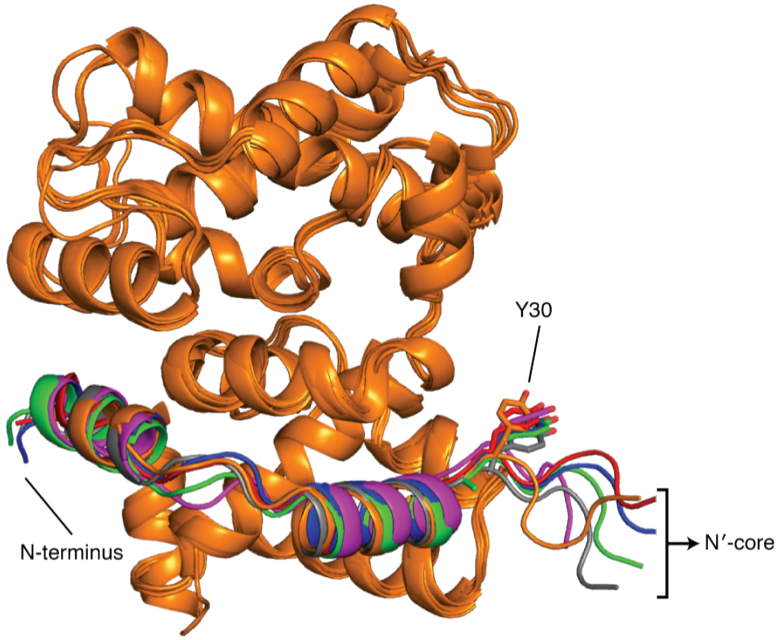


Figure 3.21 ***Superposition of N cores with the associated helical arms from neighboring subunits.***
The N cores are rendered in ribbon form and colored orange. The helical arms include residues 5-30 and are colored by structure. Orange, RVFV N_6-RNA_{35}; green, RVFV N_5-RNA_{35}; blue, RVFV N_4-RNA_{28}; gray; RVFV N_6, magenta, TOSV N_6; red, RVFV N_6-DNA_{30}.

The phlebovirus RNPs have the smallest N proteins and represent the extreme of asymmetry among a range of architectures displayed by RNPs of negative-sense viruses. The arenaviruses also appear to have an asymmetric RNP (33, 155), but the N protein has an additional exoribonuclease domain, and is unrelated to the phlebovirus N (153, 156, 157). The RNP architecture is understood in greatest detail for those viruses with RNP of highest symmetry. Respiratory syncytial virus (RSV), vesicular stomatitis virus (VSV) and rabies virus have highly symmetric, helical RNPs whose structure determines the morphology of the rod- or bullet-shaped virus particles (93-95, 102). In contrast, spherical phlebovirus particles have T=12

icosahedral symmetry that is not determined by the shape of RNP (158). Crystal structures of N-RNA multimers from RSV, VSV and rabies virus have a strict 10- or 11-fold circular symmetry that is clearly related to the helical symmetry of their respective RNPs (93-95). The related N proteins from these viruses and from Borna disease virus (96) have protruding chain termini that mediate subunit contacts in addition to substantial core-core contacts. The protrusions are in fixed positions and not flexibly hinged to the N core domain, in contrast to the helical arm of RVFV N. The RNP of influenza virus may have an intermediate degree of symmetry. The influenza N protein can occur in multiple associated states (33, 92, 97, 159), like phlebovirus N, but the influenza RNP has a clearly helical structure (33, 159, 160).

The structure of RVFV N bound to RNA enables calculation of the quantity of encapsidated genetic material that can pack into a virus particle, an important question for viruses with segmented genomes such as the phleboviruses, arenaviruses and the influenza viruses. The core of the RVFV particle has an average volume of ~150x10^6 Å^3 (158). The volume of a complete tripartite encapsidated genome is ~61x10^6 Å^3 (11,980 nucleotides, one N per 7 nucleotides, and one RNA-dependent RNA polymerase molecule per genome segment) (161). Therefore, more than two complete genomes could fit into a virus particle if packed at maximum density. Dense packing is consistent with experimental results that suggested RVFV particles contain an average molar ratio of 1:4:4 for the large (L):medium (M):small (S) genomic segments (total of nearly 29,000 nucleotides., with a packing volume of ~5200 Å^3 per nucleotide) (162). However, other negative-sense viruses for which internal volumes and RNP structures are established pack at a much lower density. If the virus particle enclosed one tripartite genome, RVFV would have a more typical packing volume of ~13,000 Å^3 per nucleotide, compared to ~21,000 Å^3 for RSV, ~17,000 Å^3 for influenza virus, and ~72,000 Å^3 for VSV (94, 102, 160).

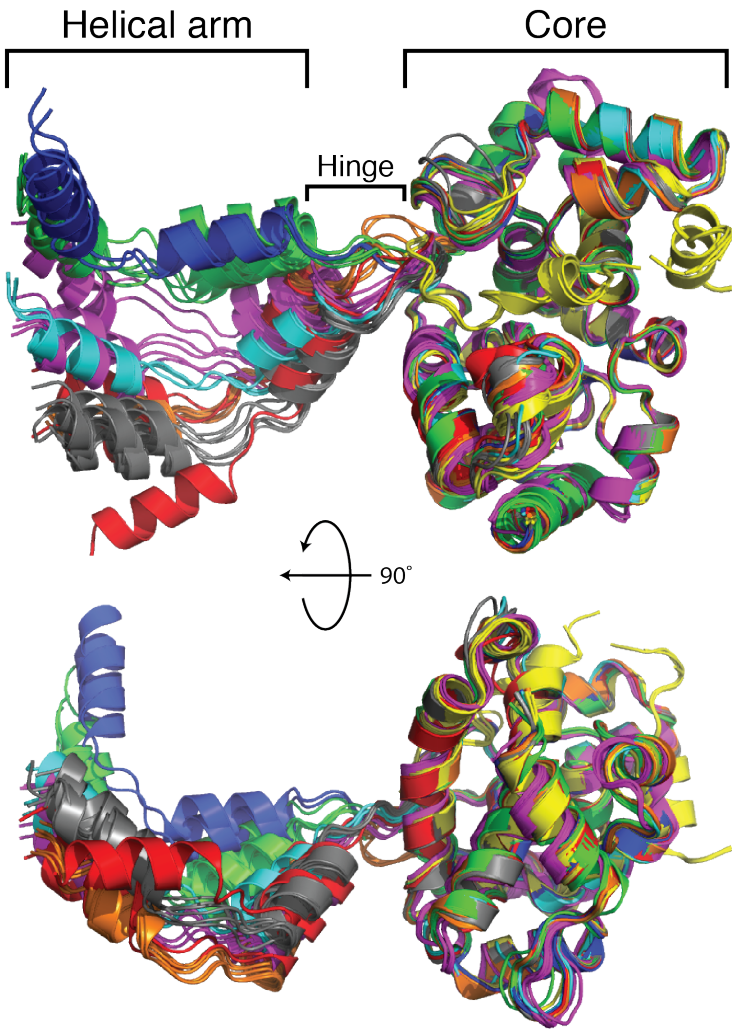


Figure 3.22 ***Flexible hinge between the helical arm and N core.***
The cores of all N subunits (residues 40-245) are superimposed. N subunits are rendered in ribbon form and colored by structure. Orange, RVFV N_6-RNA_{35}; green, RVFV N_5-RNA_{35}; blue, RVFV N_4-RNA_{28}; gray, RVFV N_6; cyan, RVFV N_6 (3OV9 (146)); yellow, RVFV N monomer (3LYF (145)); magenta, TOSV N_6; and red, RVFV N_6-DNA_{30}.

The crystal structures and EM visualization demonstrate the inherent ability of phlebovirus N to self-assemble in a simple linear or circular form (Figs 3.11-3.13). Circular RNPs have been visualized by EM for several segmented negative-sense viruses, including RVFV, influenza virus, the arenavirus Pichinde, the bunyavirus La Crosse and the phlebovirus Uukuniemi (33, 99, 103, 145, 155, 163). These viruses have short complementary sequences (12 nucleotides in RVFV) at the 5′ and 3′ ends of each genomic segment, which can create a base-paired "panhandle" structure. Therefore, phlebovirus RNP may form circles in two ways, one by RNA panhandles

and the other by N-N interactions. In influenza virus, the base-paired panhandle in each genome segment serves as a platform for RdRp attachment to the RNP and as a promoter (159). Bunyavirus transcription and replication require both N and the panhandle sequences (88, 162, 164, 165), and preparations of RVFV RNP contain both N and RdRp (145). A base-paired panhandle could function as an RdRp attachment site and promoter for each genomic segment of phleboviruses and other bunyaviruses, similar to the influenza panhandle. This is consistent with the exclusively single-stranded N-RNA interaction observed in the crystal structures, which demonstrate that N must be stripped from the RNP for RdRp "reading" during transcription and replication.

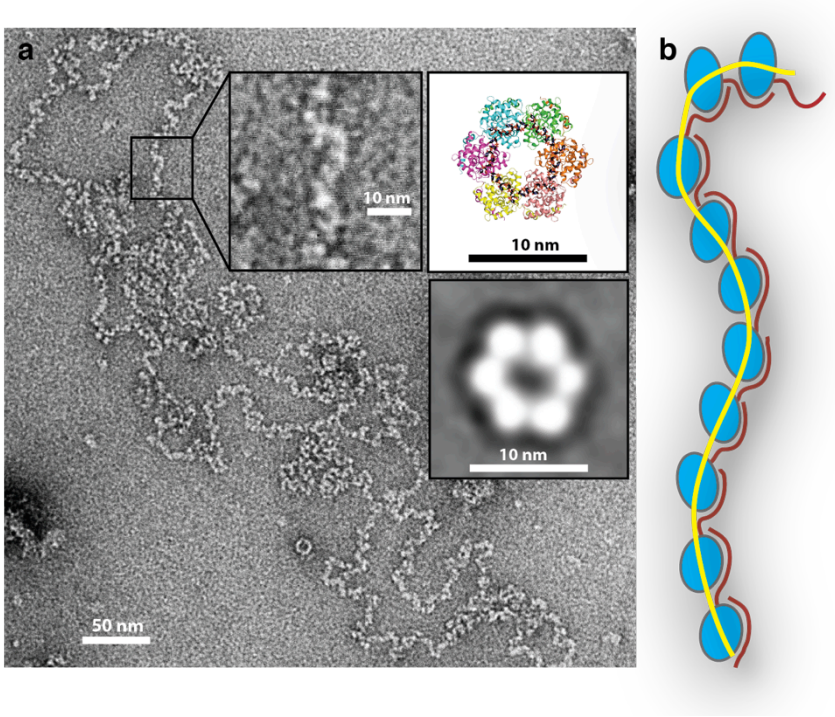


Figure 3.23 ***Negative-stain EM visualization of RVFV RNP isolated from infected BSR-T7/5 cells.*** The insets show an enlarged section of the viral RNP, the structure of the N_6-RNA_{35} hexamer (Fig. 3.11) and a predominant EM class average of N_6-DNA_{35} (Fig. 3.9c). **b**. Cartoon of RNP constructed from RNA-N_{core}-arm building blocks. The orientation of the helical arm (red arc) relative to the N core (blue oval) reflects the range of motion visualized in the crystal structures. RNA is represented as a yellow line. Authentic RVFV RNP was prepared by Mary Piper and Sonja Gerrard.

The crystal structures of RVFV N bound to RNA and EM visualization of viral and reconstituted RNPs provide a detailed understanding of how phleboviruses encapsidate their RNA genomes. The virus uses a primitive but effective system

employing limited protein-protein interactions and a deep RNA-binding slot where all RNA bases are inaccessible to other proteins or nucleic acids. The resulting nuclease-resistant RNP lacks symmetry, and becomes the template for genome replication and transcription by the RVFV RNA-dependent RNA polymerase (RdRp). Shielding of the viral RNA bases in the RNA-binding slot prevents base pairing and protects the RNA from the cellular antiviral response; however, it requires that the RdRp must strip the genome of N to access the genomic information during replication and transcription.

CHAPTER 4

STRUCTURE OF LANGAT VIRUS METHYLTRANSFERASE

Summary

Flaviviruses are an emerging public health threat due to the global distribution of their vectors and the lack of antiviral therapeutics or vaccines. Flavivirus non-structural protein 3 (NS3) and non-structural protein 5 (NS5) constitute the core of the flavivirus replicase complex, which is responsible for replicating the genome, capping the 5′ end and methylating the N7 of the guanyl cap and the 2′-O of the first base. NS3 is a multifunctional protein with an N-terminal serine protease domain and a C-terminal RNA helicase domain, while NS5, which is also a multi-domain protein, has an N-terminal methyltransferase (MTase) domain and a C-terminal RNA-dependent RNA-polymerase (RdRp) domain.

Although structures of many mosquito-borne flavivirus MTases are available, the MTase domains of tick-borne flaviviruses have not been investigated. Here we report high-resolution structures of Langat virus methyltransferase (LVMT) in apo form and in complex with SAH, GTP and GpppA cap analogues. The structures provide a detailed view of GTP and RNA cap binding to the MTase GTP-binding site and suggest possible interactions that confer guanine specificity. An RNA-cap bound structure provides the first view of an RNA cap fully extended in the active site cleft and shows the correct orientation of the RNA cap prior to ribose 2′-O methylation at the first nucleotide. Biochemical characterization of the LVMT guanylyltransferase activity using radiolabeled showed the formation of the MTase-GMP adduct, which is required for the capping reaction. Taken together, these data enhance our understanding of the cap formation process in flaviviruses.

Introduction

Flaviviruses constitute an important class of human pathogens due to the global distribution of the arthropod vectors and the high rates of human infection (34, 166). The life cycle of arthropod-borne flaviviruses involves complex relationships among insect vectors, vertebrate reservoirs, humans, and the environment (167). The spread of West Nile virus (WNV) and the dengue fever virus (DENV) illustrates the adaptability and extensive range of flaviviruses. Since 1999, WNV has caused thousands of human infections and hundreds of deaths across the continental United States, southern Canada and northern Mexico (168). DENV is endemic to more than 110 countries and infects more than 100 million people worldwide each year (169, 170). DENV infections hospitalize 500,000 individuals annually, with 12,500-25,000 suffering from the severe disease manifestations of dengue hemorrhagic fever or dengue shock syndrome (169-171). Despite the prevalence of WNV and DENV, there are no effective therapeutics to treat infections (11).

While DENV, WNV, and yellow fever virus (YFV) are mosquito-borne flaviviruses, other important pathogenic flaviviruses are transmitted by ticks. Tick-borne flaviviruses are the causative agent for the most dangerous neuroinfections in Asia and Europe (4). These viruses are cross-reactive in serological test and are known to cause encephalitic disease in humans (4). The most important tick-borne flaviviruses are tick-borne encephalitis virus (TBEV), Omsk hemorrhagic fever virus (OHFV), Powassan virus (POWV) and Langat virus (LGTV) (4). LGTV was first isolated in Malaysia and Thailand from ticks in forested areas. In those forested areas, LGTV infects rodents but does not cause overt disease (172). Remarkably, no registered cases of LGTV infection have been registered, although antibodies against the virus have been detected in the serum of local inhabitants of areas with infected ticks (4, 173, 174). The lack of overt disease in people infected with LGTV indicates that the virus is naturally attenuated.

All flaviviruses have an ~11 kB single-stranded positive-sense RNA genome, which is translated in one open-reading frame by the host cells following infection (39-41).

The polypeptide is processed by host and viral proteases to produce three structural proteins and seven non-structural proteins (42-44, 175). Non-structural protein 3 (NS3) is a multifunctional protein with helicase and protease activity. The N-terminal domain of NS3 (residues 1-170) and its cofactor NS2B associate to form a serine protease, while the C-terminal domain (residues 180-625) has RNA triphosphatase activity and structurally conserved motifs found in the DEXH family of RNA helicases including the Walker A and Walker B motifs (48-51, 53, 54). NS5 contains two domains with sequence and structural homology to *S*-adenosyl-L-methionine (SAM) -dependent methytransferases (MTase) and RNA-dependent RNA polymerases (RdRp) (63, 64). NS3 and NS5 associate to create a replicase complex, which is responsible for replicating the genome, for capping the 5′-end of newly synthesized genomic RNA and for methylating the RNA cap (65, 66, 75-77, 176).

In eukaroyotes, the 5′ RNA cap protects mRNA from nucleolytic digestion and acts as a tag for ribosome recognition (80). In flaviviruses, the viral replicase complex caps the 5′ end of the genomic RNA to mimic mature cellular mRNA (Fig. 4.1a) (42). Flaviviruses produce a mature 5′ RNA cap in four steps. First, the helicase domain of NS3 removes the γ phosphate from the 5′ end of the newly synthesized RNA genome (177). Next, a GMP moiety from GTP is covalently bonded to the MTase domain and is transferred to the RNA substrate to create an inverted 5′-5′ triphosphate bridge (Fig 4.1a) (178). Finally, the MTase methylates the RNA at both the N7 position of the guanosine cap and the 2′-O of the first base using SAM as the methyl donor (Fig. 4.1b) (65, 66, 176).

Phylogenetic analysis of NS5 proteins revealed that flaviviruses cluster into three major clades: tick-borne flaviviruses, mosquito-borne flaviviruses and flaviviruses with no known vectors (179, 180). The three clades evolved due to geographic isolation and the relationship between the invertebrate vectors and their vertebrate host (179, 180).Mosquito- and tick-borne flavivirus RNA genomes are 40-50% identical yet they exhibit differences in pathology ranging from no overt disease to encephalitis, immunity suppression and hemorrhagic diseases (4, 181). Although structures of MTase domains from mosquito-borne flaviviruses are available, the

MTase domain of a tick-borne flavivirus has not been investigated. To examine possible differences in flavivirus MTase structure and function between mosquito-borne and tick-borne viruses, we performed structural and biochemical analysis of Langat virus MTase (LVMT) cap-binding and guanyltransferase properties. Unlike other flavivirus MTase crystal structures, the RNA cap bound LVMT shows the correct orientation of the RNA substrate for ribose 2′-O methyltransfer. We also demonstrated formation of the covalent MTase-GMP adduct with a guanylyltransfers assay using radiolabeled GTP.

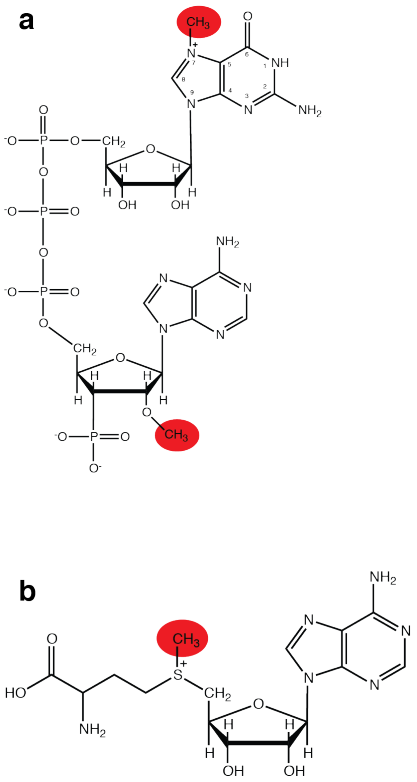


Figure 4.1 ***Flavivirus 5′ cap methylation*****.**
(a) The cap and the first base of the flavivirus genome are connected through an inverted 5′-5′ triphosphate bridge. The methyl groups added by the flavivirus methyltransferase are highlighted in red. (b) Structure of *S*-adenosyl-L-methionine (SAM) with the donated methyl group highlighted in red.

Experimental Procedures

Plasmid Construction

To construct the Langat virus methyltransferase plasmid (pSUMO-LVMT), the gene portion corresponding to residues 1-270 from Langat virus NS5 was amplified using primers 5′AGATTGGTGGCggtggatccgagggaga and 5′GAGGAGAGTTTAGACATTAct-ctgccagcactacg (IDT). The PCR product was processed with T4 DNA polymerase (Promega) and dGTP (Invitrogen) to generate overhangs for ligation-independent cloning (LIC) and then incubated with processed LIC-SUMO vector (pETHSUL), which encodes a His_6-SUMO fusion protein (147). The mixture was transformed into *E. coli* XL1-Blue competent cells for ligation and amplification of the plasmid, which was confirmed by DNA sequencing.

Protein Expression and Purification

pSUMO-LVMT was transformed into *E. coli* strain BL21 AI (Invitrogen) containing the pRARE2 plasmid (Novagen) and grown in 500 mL of TB media (6 g tryptone, 12 g yeast extract, 1.15 g KH_2PO_4(monobasic), 6.25 g K_2HPO_4 (dibasic), 20 mL glycerol) containing 35 μg/mL chloramphenicol and 100 μg/mL ampicillin at 37˚C until OD_{600}=1.0. The temperature was reduced to 20˚C, and expression was induced after 40 minutes by addition of 2 mL 50% w/v arabinose and isopropyl-β-D-thiogalactopyranose (IPTG) to a final concentration of 0.4 mM. The cultures were incubated 16 hr at 20˚C and cells were harvested by centrifugation.

All purification steps were carried out at 4˚C. Cell pellets were resuspended in 30 mL lysis buffer (25 mM Tris pH 7.8, 500 mM NaCl, 20 mM imidazole, 5% glycerol), lysed by sonication, and centrifuged at 27,000 x g for 45 minutes at 4°C. The supernatant was loaded onto a 5-mL HiTrap chelating column (GE Healthcare) pre-equilibrated with lysis buffer. The column was washed with 30 mL lysis buffer. The protein was eluted with a linear gradient of 20-500 mM imidazole in lysis buffer. Fractions containing SUMO-LVMT, as determined by 12% SDS-PAGE, were pooled and dialyzed 1 hr against 1 L dialysis buffer (50 mM Tris pH 7.8, 500 mM NaCl, 5% glycerol).

The His_6-SUMO fusion was removed by incubating SUMO-LVMT at 4°C overnight with His-tagged SUMO-hydrolase (147) at a final concentration of 1:1000 (protease:protein) and dialysis was continued with fresh buffer for 16 hr. The proteolysis mixture was loaded on a 5-mL HiTrap column pre-equilibrated with lysis buffer, and cleaved LVMT was washed from the column with lysis buffer.

LVMT was concentrated using Centriprep-10 (Millipore) and subjected to size exclusion chromatography by a HiLoad 16/60 Superdex 75 gel filtration column (Amersham) pre-equilibrated with storage buffer (20 mM Tris pH 7.8, 0.5 M NaCl, 10% glycerol). Fractions corresponding to the LVMT monomer peak were pooled and concentrated to ~10 mg/mL using Centriprep-10. Purified protein was flash-frozen in liquid N_2 and stored at -80°C. Typical 500 mL cultures yielded 10 mg of purified LVMT. Similar to other flavivirus methyltransferases, Langat virus methyltransferase was as a monomer in solution, with an A_{260}/A_{280} absorbance ratio of 0.75 (Fig. 4.2). Flavivirus methyltransferases copurify with *S*-adenosyl-L-homocysteine (SAH) or *S*-adenosyl-L-methionine (SAM), which account for the high absorbance ratio.

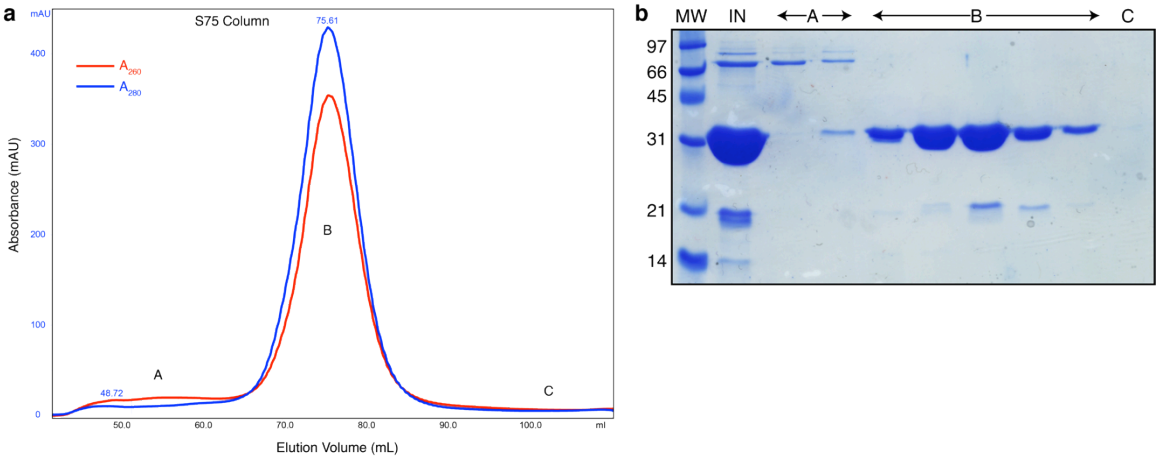


Figure 4.2 ***Purification of recombinant LVMT.***
(**a**) Preparative S75 gel-filtration chromatogram of recombinant LVMT after purification by Ni-affinity chromatography and cleavage of the SUMO fusion partner. Peak A is the void volume and peak B is LVMT. The red and blue traces represent absorption at 260 nm and 280 nm, respectively. (**b**) SDS-PAGE of fractions from the chromatogram shown in (a). Lane 1: molecular weight markers, lane 2: input sample, lanes 3-10 fractions from peaks A–C, as labeled

LVMT Crystallization

Prior to crystallization, LVMT was dialyzed against crystallization buffer (20 mM Tris pH 7.8, 200 mM NaCl). LVMT was crystallized at 20˚C by hanging drop vapor diffusion from a 1:1 mixture of protein (10 mg/mL LVMT in crystallization buffer) and well solution (22% PEG 3350, 150 mM $MgCl_2$ and 100 mM Bis-tris pH 6.5). Optimal crystals were obtained after 4 days. The crystals were cryo-protected by soaking in well solution with the addition of 15% glycerol, harvested into loops, and frozen by plunging into liquid N_2. For co-crystalization experiments, LVMT was incubated with 1 mM ^{M7}GpppA cap analogue or 5 mM guanosine-5'-triphosphate (GTP) for 1 hour prior to crystallization. For crystal soaking experiments, LVMT crystals were grown, harvested, and soaked overnight in fresh reservoir solution supplemented with 1 mM GpppA cap analogue or 5 mM GTP and 15% glycerol.

Data Collection and Structure Determination

Diffraction data were collected at 100 K on GM/CA beamline 23ID-D at the Advanced Photon Source (APS), Argonne National Laboratory (Argonne, IL) at λ = 1.0332 Å. Diffraction images were indexed and integrated with XDS (182) and scaled with XSCALE (183) (Tables 4.1-4). Dengue virus MTase (PDB accession code 1L9K) (65) was used as a model for molecular replacement with Phaser (149). Modeling was completed manually using Coot (116). Refinement was performed using Refmac (117) with translation-libration-screw (TLS) parameterization of molecular motion (152) (Table 4.5). Anisotropic thermal parameters were refined for structures with d_{min} less than 1.5 Å, and individual isotropic thermal parameters for those with d_{min} greater than 1.5 Å. Ramachandran analysis and structure validation were performed by MolProbity (119). PyMOL was used for structural alignments and to generate figures (120). The APBS plugin (122) in PyMOL was used to calculate electrostatic surface potentials and ConSurf (125) was used for calculating conservation scores. Ligand interaction plots were calculated with MOE 2009.10 (184).

Model quality and electron density

LVMT crystals contained one polypeptide in the asymmetric unit of space group $P2_12_12_1$. MolProbity analysis of the refined structures gave good statistics with scores of 1.05 for LVMT (99th percentile), 1.53 for LVMT-GpppA CAP (soaked) (98th percentile), 1.31 for LVMT-^{M7}GpppA CAP (co-crystallized) (97th percentile) and 1.23 for LVMT-GTP (soaked) (98th percentile). Ramachandran analysis showed all residues were in the allowed region except for Gly148 in LVMT-GTP and LVMT-^{M7}GpppA CAP (co-crystallized) (Figs 4.3-4.6). The refined structures are complete with the exception of residues 1-4 and 266-270. In those regions, the electron density was too weak for modeling.

Table 4.1 ***XSCALE scaling summary for LVMT.***

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR COMPARED expected		I/SIGMA
3.81	16877	2618	2668	98.1%	3.8%	4.7%	16858	37.41
2.70	30740	4549	4554	99.9%	4.5%	4.8%	30740	35.84
2.21	39806	5803	5811	99.9%	5.5%	5.3%	39806	31.21
1.91	47364	6814	6819	99.9%	7.0%	6.4%	47364	25.10
1.71	54775	7696	7700	99.9%	10.7%	10.5%	54775	16.75
1.56	60360	8433	8436	100.0%	17.2%	18.4%	60360	10.66
1.45	65428	9187	9187	100.0%	28.5%	32.1%	65428	6.65
1.35	68613	9765	9822	99.4%	49.3%	58.3%	68603	3.76
1.28	42804	7718	10455	73.8%	74.5%	89.6%	42647	2.11
total	426767	62583	65452	95.6%	6.8%	7.3%	426581	15.12

Table 4.2 ***XSCALE scaling summary for LVMT-GTP.***

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR COMPARED expected		I/SIGMA
4.77	9312	1399	1401	99.9%	2.4%	2.5%	9310	63.60
3.38	16933	2383	2384	100.0%	2.5%	2.6%	16933	63.02
2.76	21859	3021	3023	99.9%	4.0%	3.8%	21859	43.56
2.39	25669	3524	3524	100.0%	6.5%	6.3%	25669	29.22
2.14	28927	3962	3963	100.0%	9.2%	9.1%	28927	21.37
1.96	32047	4393	4394	100.0%	14.7%	14.9%	32047	13.76
1.81	34456	4730	4732	100.0%	26.6%	27.2%	34456	7.85
1.69	36903	5082	5083	100.0%	45.1%	46.5%	36902	4.64
1.60	37799	5292	5363	98.7%	68.8%	71.0%	37759	2.96
total	243905	33786	33867	99.8%	7.4%	7.5%	243862	20.58

Table 4.3 ***XSCALE scaling summary for LVMT-GpppA (soaked).***

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR expected	COMPARED	I/SIGMA
6.32	3930	613	614	99.8%	2.7%	3.0%	3930	48.88
4.49	7134	1023	1023	100.0%	3.5%	3.8%	7134	42.58
3.67	9148	1289	1290	99.9%	4.1%	4.0%	9148	40.56
3.18	10823	1503	1504	99.9%	6.5%	6.3%	10823	28.18
2.85	12245	1692	1692	100.0%	13.2%	13.1%	12245	15.49
2.60	13454	1848	1850	99.9%	22.2%	22.4%	13454	9.67
2.41	14719	2023	2023	100.0%	34.1%	34.4%	14719	6.64
2.25	15676	2146	2146	100.0%	49.4%	50.2%	15676	4.76
2.13	12705	1999	2272	88.0%	67.4%	67.1%	12616	3.17
total	99834	14136	14414	98.1%	11.8%	11.9%	99745	17.13

Table 4.4 ***XSCALE scaling summary for LVMT-^{M7}GpppA (co-crystallized).***

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR expected	COMPARED	I/SIGMA
5.17	7179	1115	1119	99.6%	3.4%	3.9%	7178	41.61
3.67	12726	1873	1874	99.9%	3.7%	3.9%	12726	41.96
3.00	16728	2386	2386	100.0%	4.9%	4.7%	16728	34.15
2.60	19895	2771	2771	100.0%	7.3%	7.0%	19895	24.50
2.33	22604	3119	3119	100.0%	10.4%	10.1%	22604	18.43
2.13	25101	3451	3451	100.0%	15.0%	15.2%	25101	13.25
1.97	27204	3728	3729	100.0%	23.7%	25.0%	27204	8.71
1.84	28938	3976	3977	100.0%	40.5%	44.2%	28938	5.11
1.74	26454	3977	4231	94.0%	64.9%	71.5%	26353	2.93
total	186829	26396	26657	99.0%	8.3%	8.5%	186727	16.75

LVMT GTPase assay

LVMT was dialyzed overnight in a buffer containing 20 mM Tris pH 7.8 and 100 mM NaCl. The assay was performed by incubating 1μM, 2 μM and 3 μM of LVMT with 1 mM [α-^{32}P]GTP (PerkinElmer) in a buffer containing 20 mM Tris pH 7.8, 100 mM NaCl, 5 mM DTT, and 5 mM $MgCl_2$ at 30°C for 1 hour. The reactions were stopped by the addition of SDS-PAGE running buffer containing 1% SDS. Reaction samples were analyzed by electrophoresis with a 12% polyacrylamide gel containing 0.1% SDS. Radiolabeled proteins were visualized by autoradiography of the gel (178).

Table 4.5 ***Crystallographic Summary.***

	LVMT apo	LVMT GTP	LVMT CAP (soaked)	LVMT CAP (Co-crystallized)
Diffraction data				
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell lengths (Å) a, b, c	44.2 62.3 89.9	44.5 62.24 90.0	43.9 62.1 89.8	44.4 62.4 90.4
X-ray source	APS 23 ID-D	APS 23 ID-D	APS 23 ID-D	APS 23 ID-D
Wavelength (Å)	1.0332	1.0332	1.0332	1.0332
d_{min} (Å)	1.28 (1.35-1.28)	1.60 (1.69-1.60)	2.13 (2.25-2.13)	1.74 (1.84-1.74)
Unique reflections	62,583	33,789	33,786	26,396
R_{merge}[b]	0.06 (0.75)	0.07 (0.68)	0.12 (.67)	0.08 (0.64)
Avg I/σ_I	15.1 (2.1)	20.6 (3.0)	17.13 (3.17)	16.75 (2.93)
Completeness (%)	95.6 (73.8)	99.8 (98.7)	98.1 (88.0)	99.0 (94.0)
Average redundancy	6.8 (5.5)	7.2 (7.1)	7.0 (6.3)	7.0 (6.6)
Refinement				
Data range (Å)	39.68-1.28	39.90-1.60	44.91-2.13	39.87-1.74
Reflections	59,452	32,095	13,428	25,075
R/R_{free}[c]	0.153/0.185	0.160/0.203	0.183/0.224	0.186/0.222
RMS deviations				
Bonds (Å)	0.010	0.006	0.008	0.009
Angles (°)	1.45	1.18	1.26	1.32
Avg B-factor (Å^2)				
Protein	14.8	16.0	21.7	18.5
Water	26.5	25.8	31.1	29.7
Ligands	12.2	21.3	38.8	29.7
Ramachandran[d]				
Allowed (%)	100	99.6	100	99.6
Outlier (%)	0	0.4	0	0.4
Number of Atoms				
Protein	2062	2062	2062	2062
Water	244	219	217	229
Ligands	26	59	77	78

[a] Values in parentheses are for the outermost shell of data.

[b] Rmerge = $\sum|Ii-\langle I\rangle|/\sum Ii$, where Ii is the intensity of the ith observation and <I> is the mean intensity.

[c] $R = \sum||F_o|-|F_c||/\sum|F_o|$. R_{free} is calculated for a 5% subset of the data

[d] Calculated with MolProbity (119).

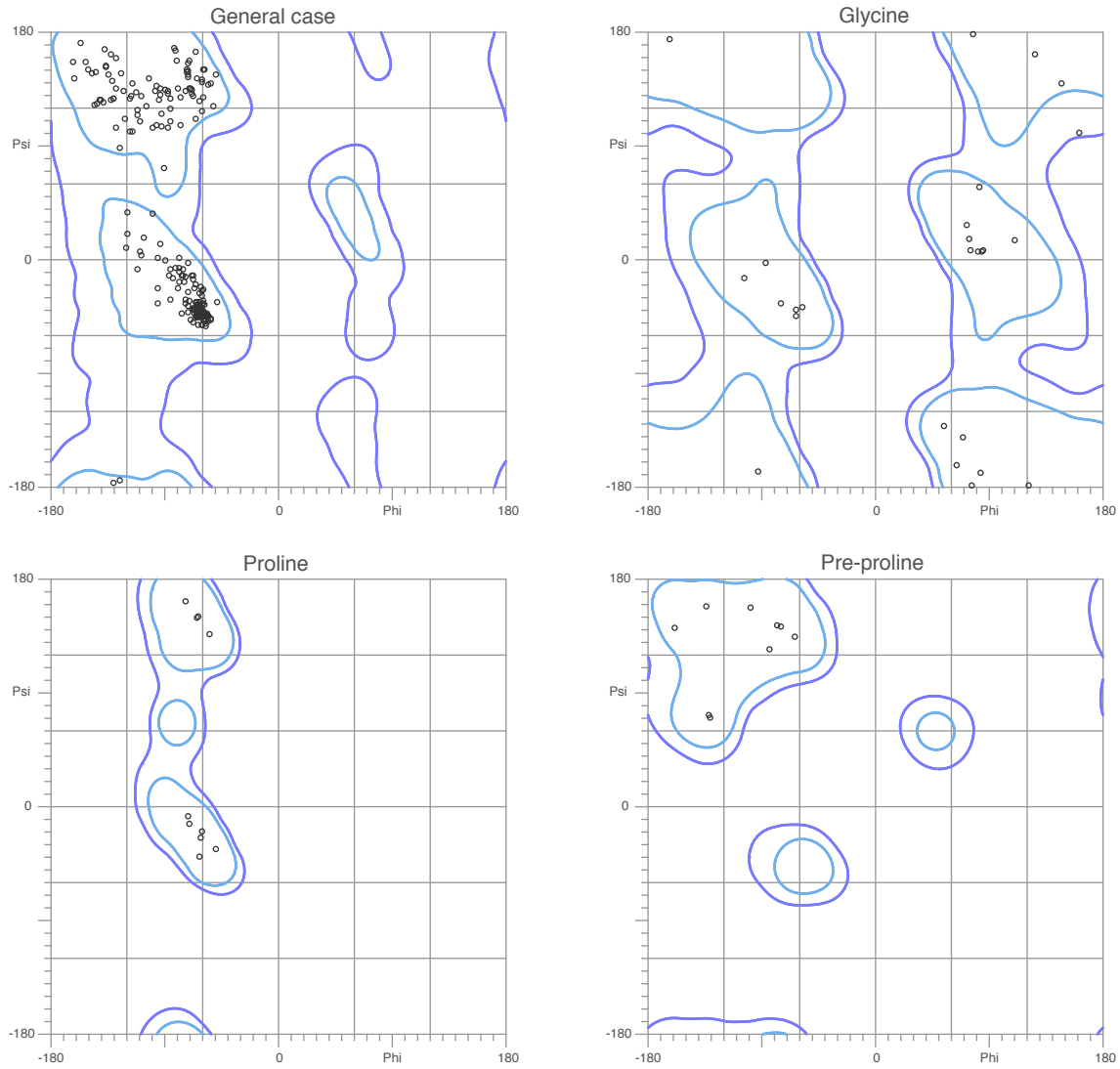


Figure 4.3 ***Ramachandran plot for LVMT.***
As output from MolProbity, 100% of residues were in the allowed regions.

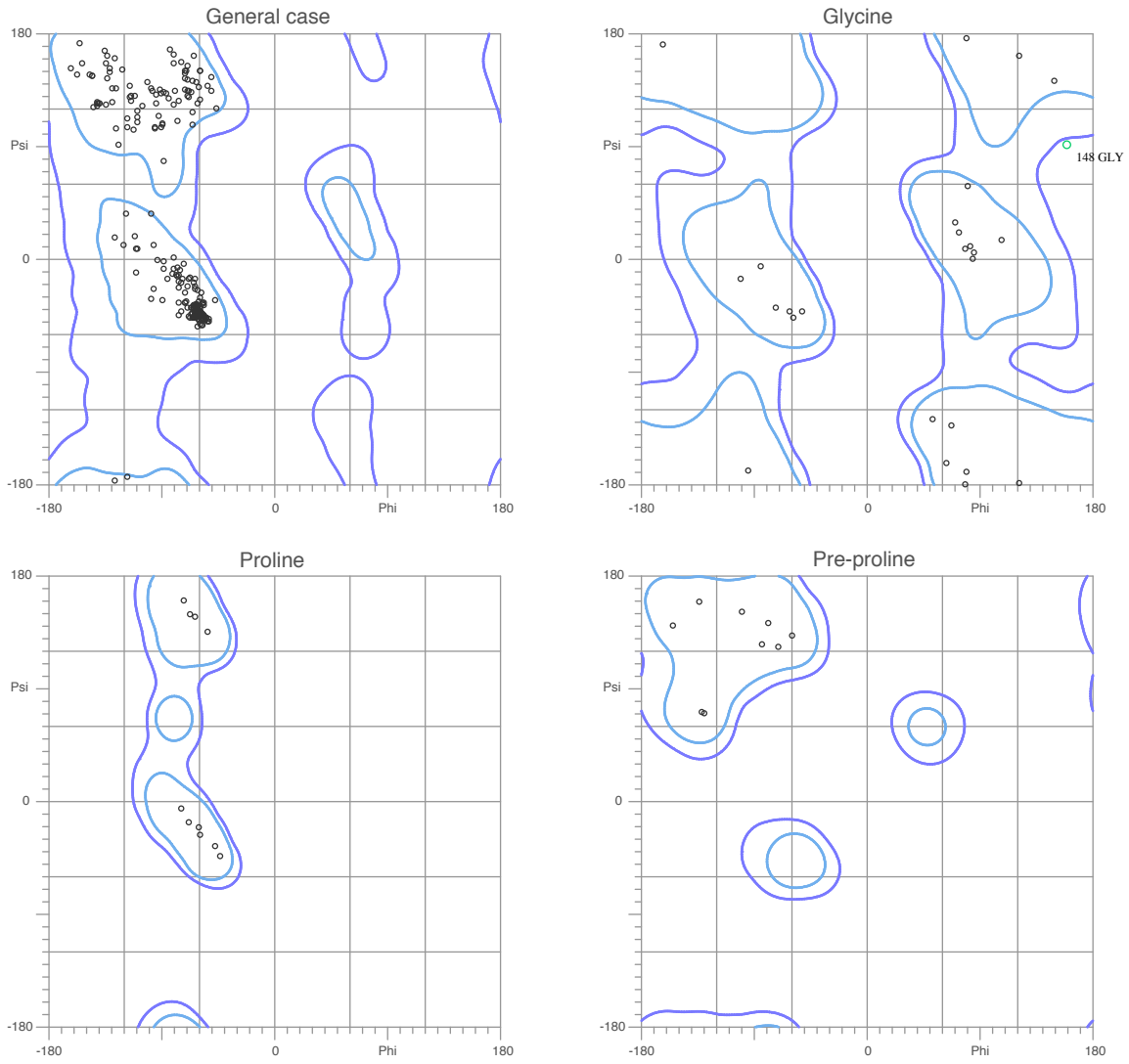


Figure 4.4 ***Ramachandran plot for LVMT-GTP.***
As output from MolProbity, 99.6% of residues were in the allowed regions.

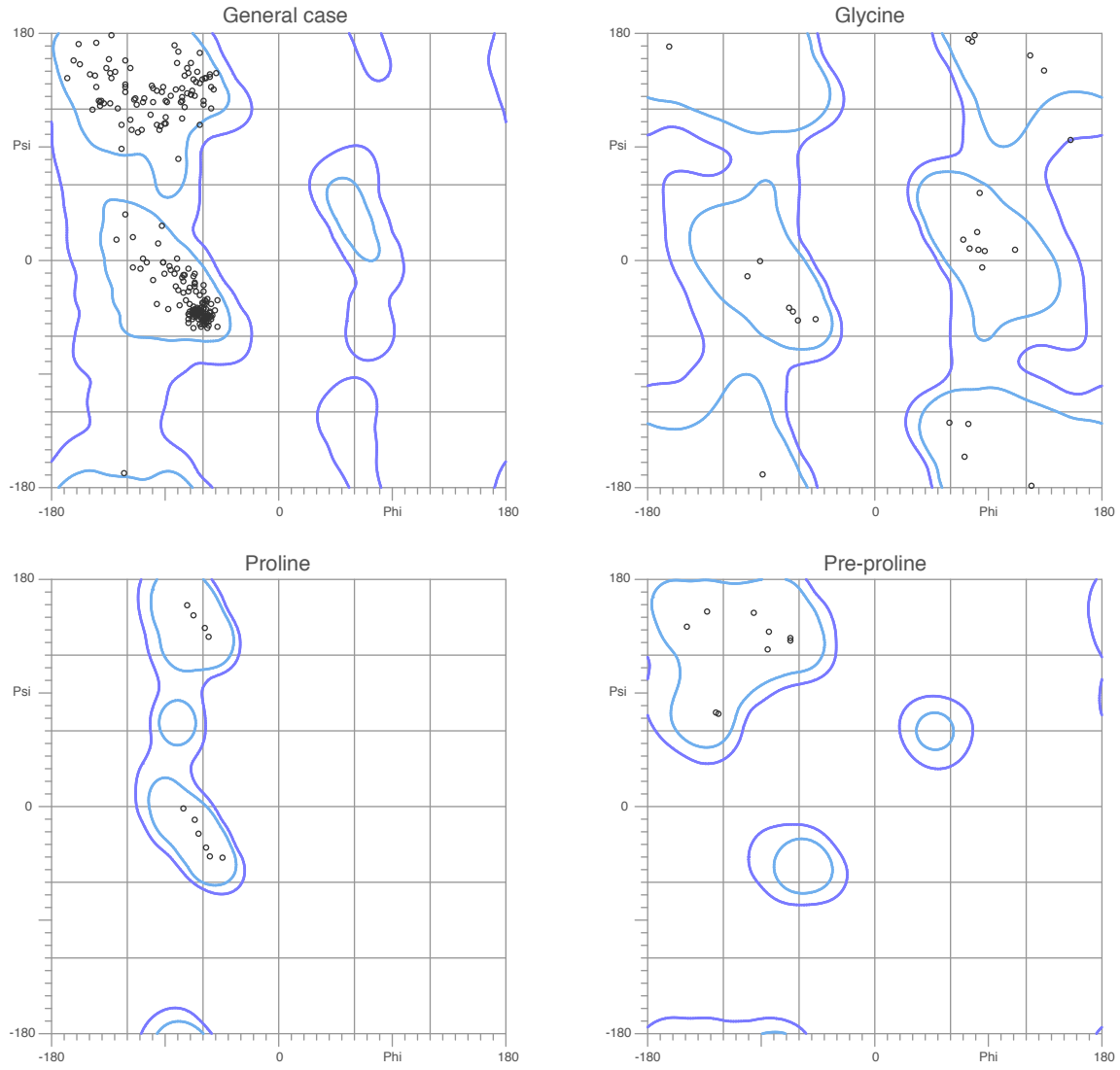


Figure 4.5 ***Ramachandran plot for LVMT-GpppA (soaked).***
As output from MolProbity, 100% of residues were in the allowed regions.

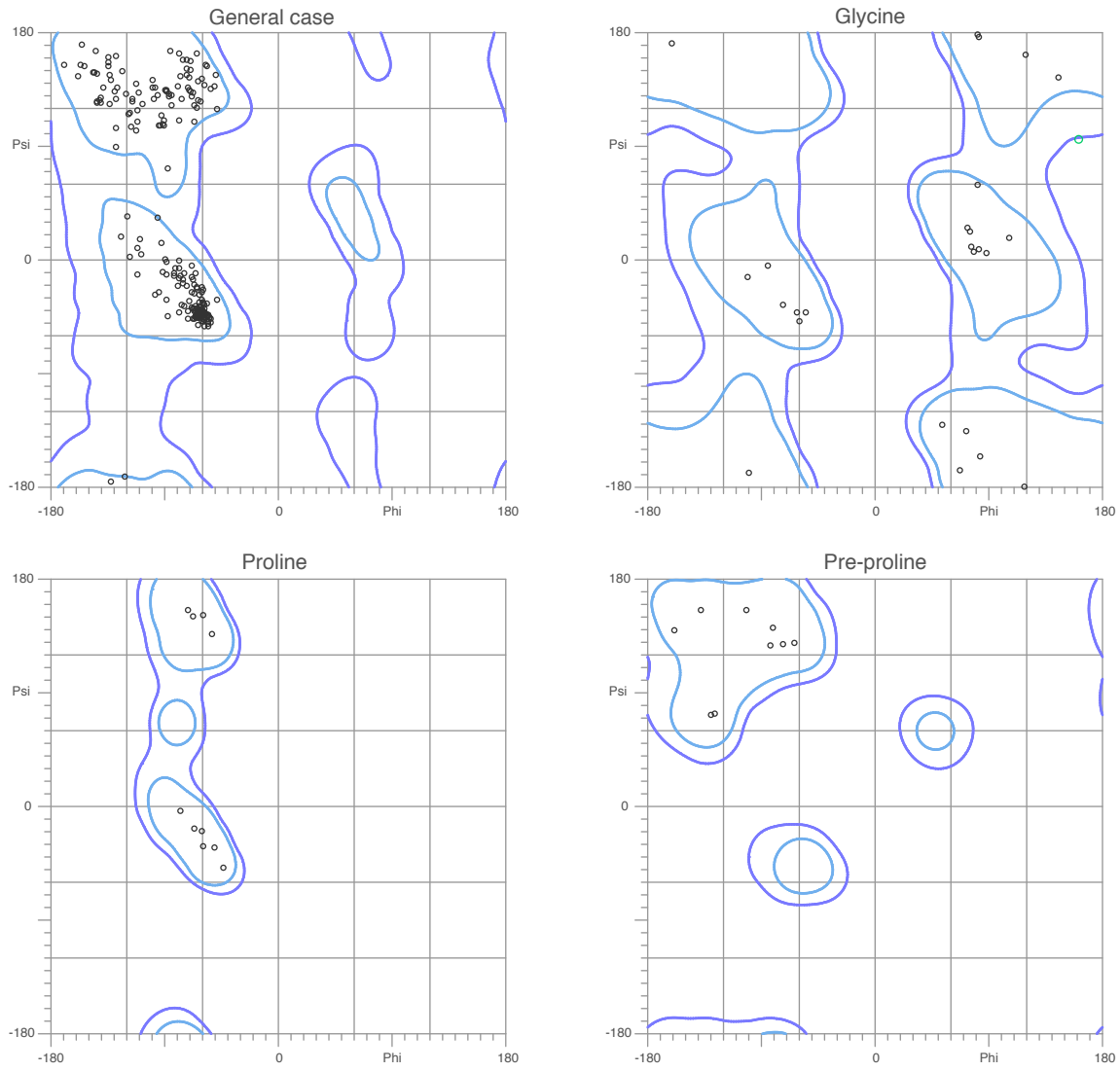


Figure 4.6 ***Ramachandran plot for LVMT-^{M7}GpppA (co-crystallized).***
As output from MolProbity, 99.6% of residues were in the allowed regions.

Results

Structure of LVMT

The Langat virus methyltransferase domain has a compact globular structure with approximate dimensions of 55-Å X 42-Å X 40-Å (Fig. 4.7a). The core of the protein has seven β strands arranged as a parallel β sheet (Fig. 4.7b), similar to the catalytic domains of other class I Rossmann-fold-like SAM-dependent methyltransferases (185). The β sheet is surrounded by eight α-helices and resembles the α-β-α sandwich fold seen in nucleotide-binding proteins (Fig. 4.7b). A wide cleft in the MTase domain is the site of ligand binding, methyltransfer and possibly guanyltransfer. Strong electron density was observed for *S*-adenosyl-L-homocysteine (SAH) in all structures (Fig. 4.8). SAH bound in a back pocket between β2 and β3 of the core β sheet (Figs 4.8 & 4.9). As expected due to the high sequence identity, the Langat virus methyltransferase domain is similar to those from other flavivirus MTases (average RMSD= 0.75 Å over 250 residues for 4 structure cmparisons) (Figs 4.10 & 4.11) (65, 176, 186). The only exceptions were slight conformational changes in the surface loops (Fig. 4.10).

GTP binding

Flavivirus RNA is methylated at the N7 position of the GTP cap and the 2'-O of the invariant 5'-adenosine nucleotide (65, 176). The flavivirus MTase is responsible for methylation at both positions. To investigate Langat virus MTase GTP and cap binding, we performed soaking and co-crystallization experiments with recombinant LVMT and GpppA cap analogues and GTP. In all cases, the crystals showed clear positive difference density for the ligands in a GTP binding site, which is 17 Å from the SAM binding site (Fig. 4.12a). Strong difference density was observed for the guanine and ribose moieties in the GTP-binding site (Fig. 4.12a). Positive difference density corresponding to the phosphate groups extended towards the SAH binding site and was strongest for the α and β phosphate groups, with weak or no electron density for the γ-phosphate. As seen in other flavivirus GTP or cap analogues bound MTase structures, GTP binding is mediated by π-π

stacking interactions with invariant Phe24 and a hydrogen-bonding network interacting with the ribose and phosphate groups (Fig. 4.13). There are few interactions with the guanine base (Fig. 4.13). Two specific backbone interactions with the guanine base by Leu16 and Cys19 (both invariant in flavivirus MTases) may determine GTP specificity (Fig. 4.13).

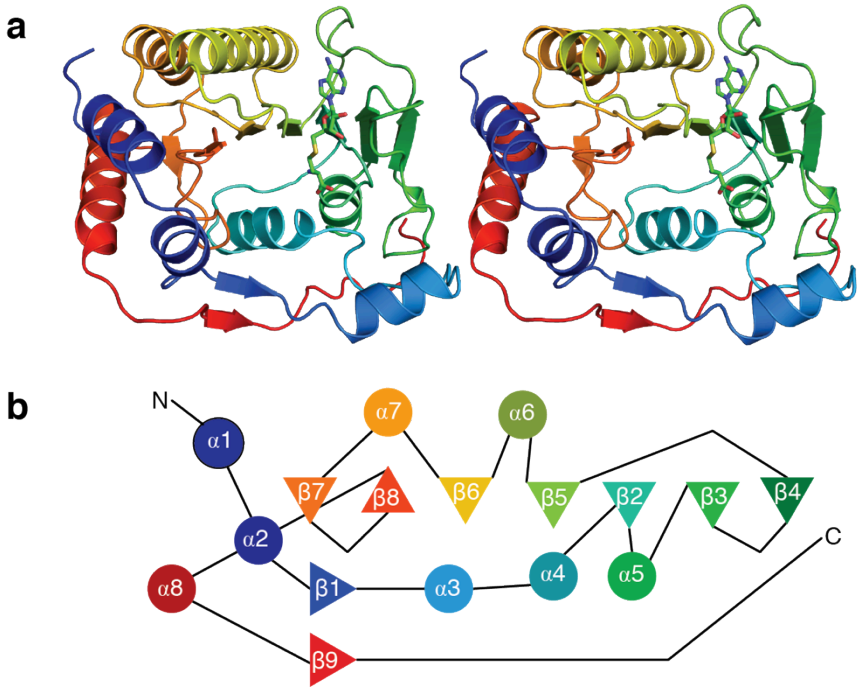


Figure 4.7 ***Structure of LVMT.***
(**a**) The stereo ribbon diagram is colored as a rainbow from blue at the N- terminus to red at the C-terminus with SAH in stick form with green carbon atoms. (**b**) Topology diagram of LVMT showing the ado-Met-dependent MTase fold. Circles and triangles represent α-helices and β-strands, respectively, and are colored as in (**a**) with black lines representing loops.

GpppA cap analogue binding

Irrespective of the crystallization conditions, the guanosine in the GpppA cap analogue binds identically to the guanosine in GTP-bound structures (Fig 4.12) (66, 186). However, the conformation of the triphosphate and adenine base in the cap analogues differs based on whether the cap analogues were soaked into LVMT crystals or co-crystallized with the protein. If cap analogues were soaked into the crystals, the adenine base in the cap stacks with the guanyl base in the GTP-binding site (Fig. 4.12b). The internally stacked conformation of the adenine and guanine

bases has been observed in DENV and YFV MTases crystals soaked with GpppA cap analogues (66, 186). This conformation is not biologically relevant as neither the N7 of the guanine cap nor the ribose 2'-O of the adenyl base are positioned close enough to the SAM-binding site for methyl transfer (Fig. 4.12b). In contrast, when cap analogues were co-crystallized with the LVMT, the base stacking between the two bases in the cap is not observed (Fig. 4.12c). Strong positive difference density for the guanosine is observed in the GTP-binding site along with the α and β phosphate groups (Fig. 4.12c). The phosphate groups extend towards the SAM-binding site similar to the GTP-soaked crystals. There was weak, diffuse electron density for the adenosine and the ribose moiety of the adenine base in the space between the phosphate groups and the SAM-binding site (Fig. 4.12c). The density was not defined enough to confidently fit the adenosine base, suggestive of mobility (Fig 4.12c).

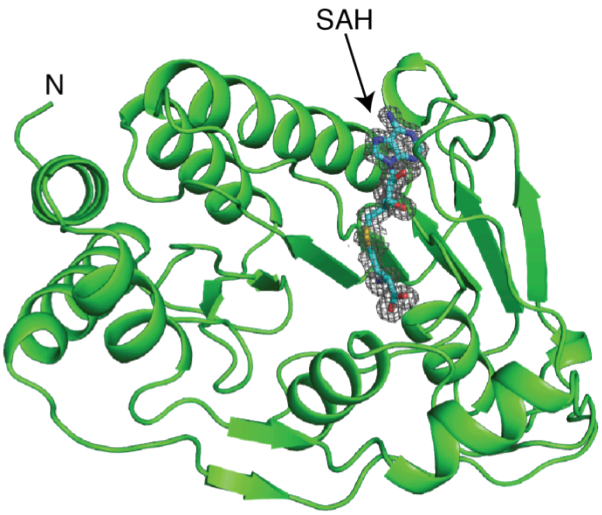


Figure 4.8 ***Flavivirus SAM-binding site*.**
The 1.28 Å Fo-Fc omit map contoured at 2.5σ of SAH bound to LVMT. LVMT is rendered a green ribbon and SAH is rendered in sticks form with cyan C atoms.

RNA binding

Flavivirus MTase must bind RNA to perform the methyltransfer reactions. An electrostatic surface potential diagram of the MTase domain showed a large positively charged surface at the entrance of the active site cleft between the GTP and SAM-binding site (Fig. 4.14a). The positively charged surface extends from the

active site cleft entrance to the GTP- and SAM-binding sites, which are conserved in flaviviruses (Fig. 4.14b). The positively charged surface is formed from nine invarient lysine residues and is likely involved in RNA binding (Figs 4.9b & 4.14c).

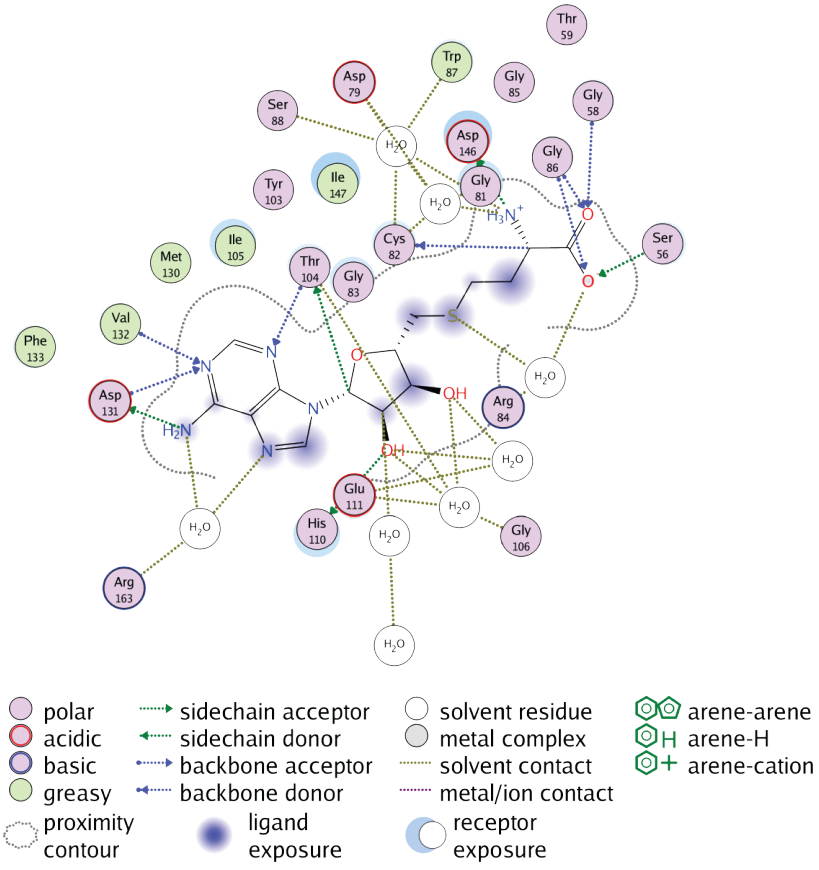


Figure 4.9 ***Ligand interaction plot of SAH in the SAM-binding site.***
A 2D representation of SAH in the SAM-binding site showing interactions with LVMT residues and ordered solvent. Figure generated with MOE 2009.10 (184).

Formation of LVMT MTase-GMP adduct

Flavivirus MTases are also guanylyltransferases capable of capping the 5′ end of RNA substrates. Biochemical studies demonstrated that DENV, YFV and WNV MTases can form a covalent MTase-GMP adduct after incubation with radiolabeled GTP in the presence of Mg^{2+} (178). We used a similar assay to show that LVMT is also capable of forming an MTase-GMP adduct. When LVMT was incubated with radiolabeled GTP and Mg^{2+}, covalent LVMT-GMP adduct formation increased with increasing protein concentration (Fig. 4.15, lane 1-3). However, adduct formation

significantly decreased when EDTA was added to the reaction mixture (Fig. 4.15, lane 4). The decrease in adduct formation supports the previously reported guanylyltransferase Mg^{2+} requirement (178). Although the crystallization conditions contained Mg^{2+}, neither the covalent adduct nor Mg^{2+} were observed in the LVMT-GTP crystal. Future experiments will further characterize this activity in flavivirus MTases.

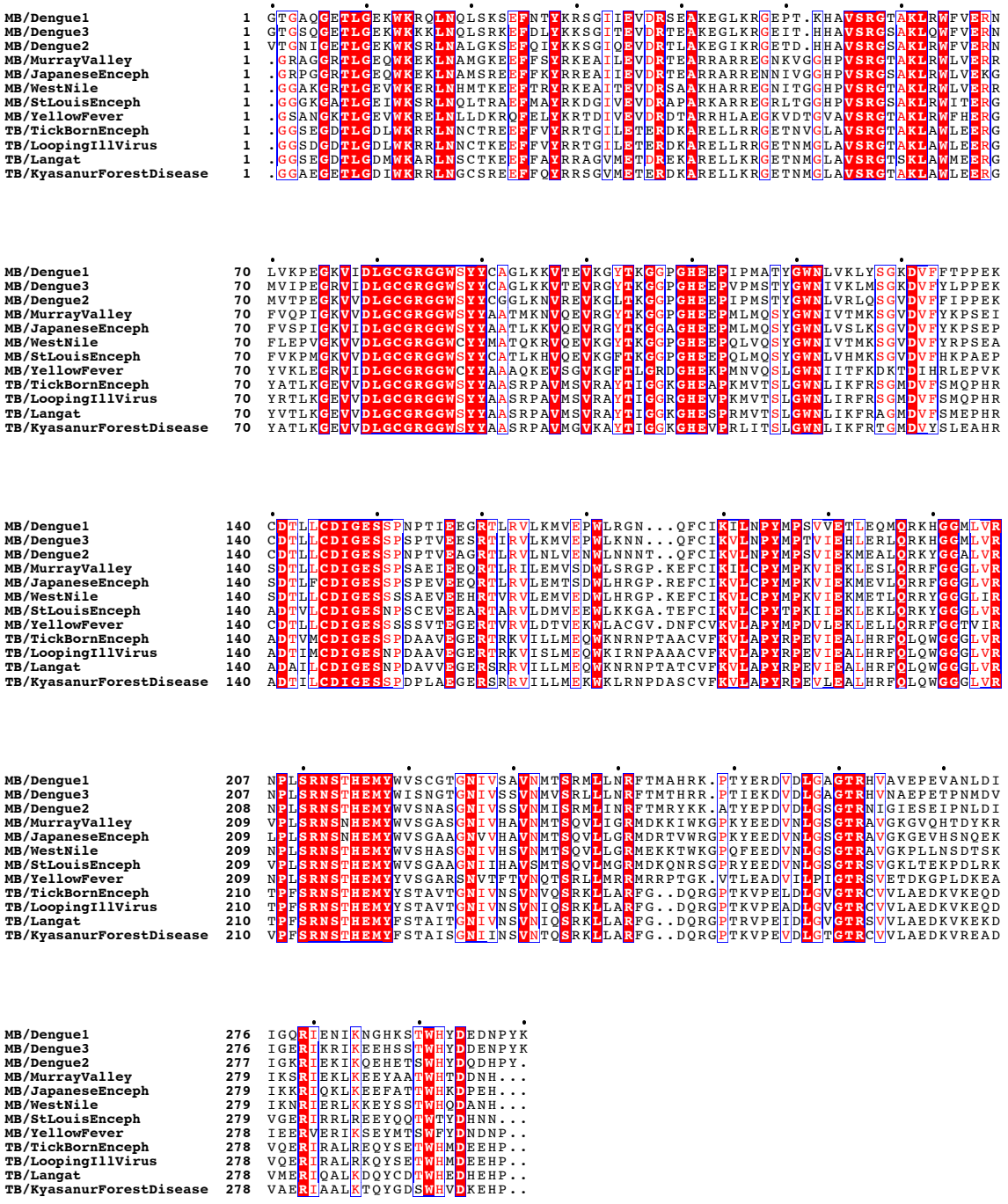


Figure 4.10 ***Sequence alignment of MTase from flaviviruses.***
Representative MTases from mosquito-borned (MB) or tick-borne (TB) flaviviruses were compared. Invariant residues are shown in white with red background, conserved residues are shown in red with while backgrounds, and variable residues are shown in black with white backgrounds.

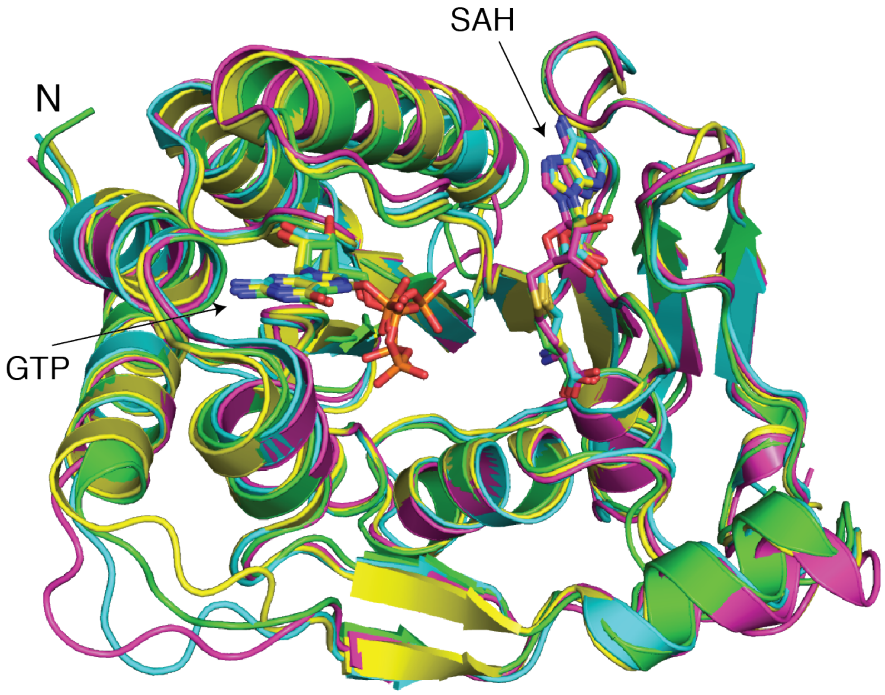


Figure 4.11 Structural conservation of flavivirus MTases domains.
LVMT, DENV, WNV and YFV MTase domains are superimposed. MTase domains are rendered in ribbon form and colored by structure. Green, LVMT; cyan, DENV; yellow, YFV; and magenta, WNV. SAH and GTP are shown rendered in stick form with C atoms colored by structure.

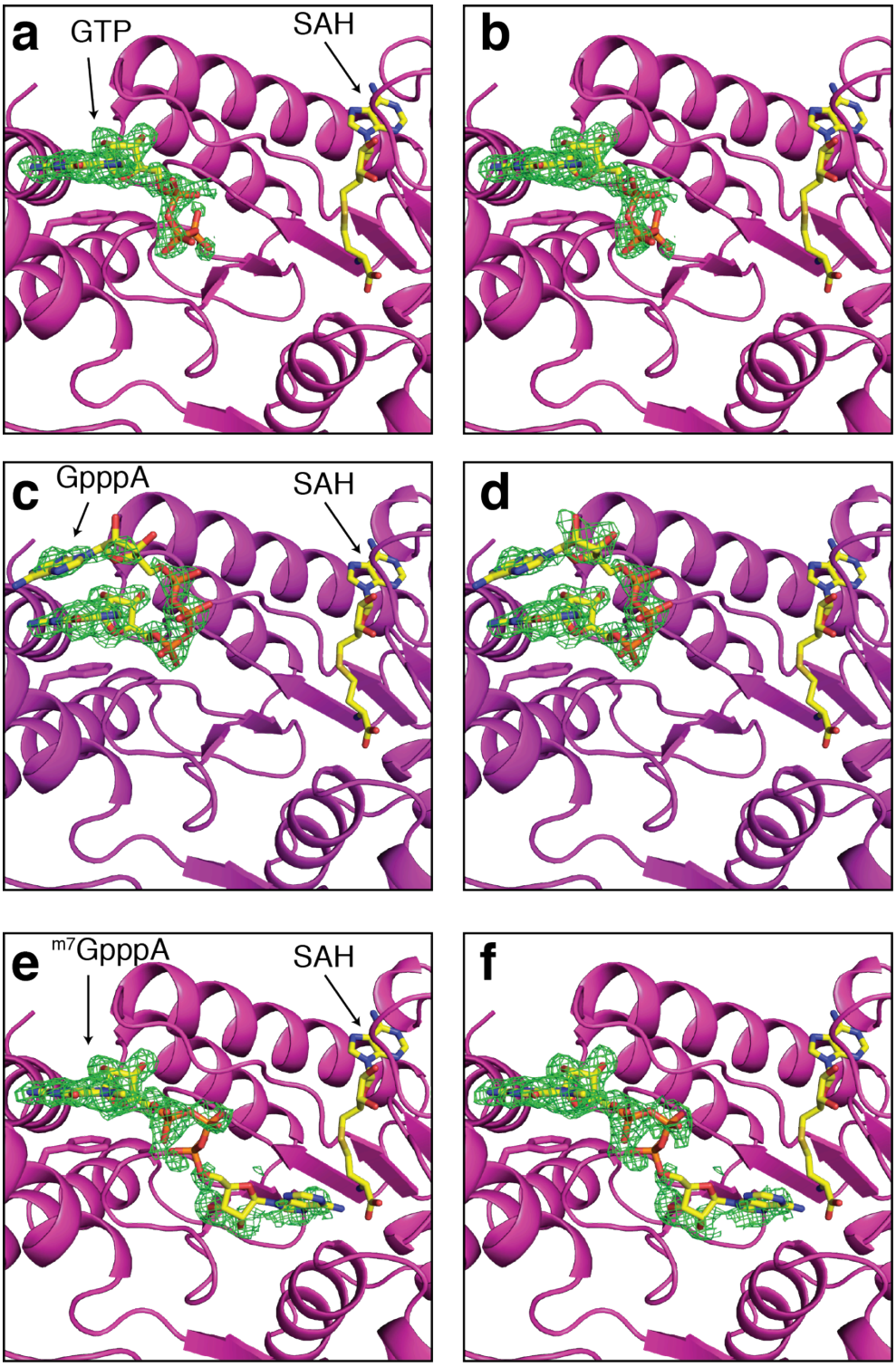


Figure 4.12 ***LVMT GTP-binding site.***
Electron density in GTP-binding site before (Fo-Fc contoured at 2.5σ) and after (omit Fo-Fc contoured at 2.5σ) addition of ligand to models of (**a**, **b**) LVMT-GTP at 1.60 Å, (**c**, **d**) LVMT-GpppA (soaked) at 2.13 Å and (**e**, **f**) LVMT-^{M7}GpppA at 1.74 Å. The MTase domains are rendered as magenta ribbons and SAH, GTP, and cap analogues are shown in sticks form with yellow C atoms. For each structure, the left and right panels show original and omit density, respectively.

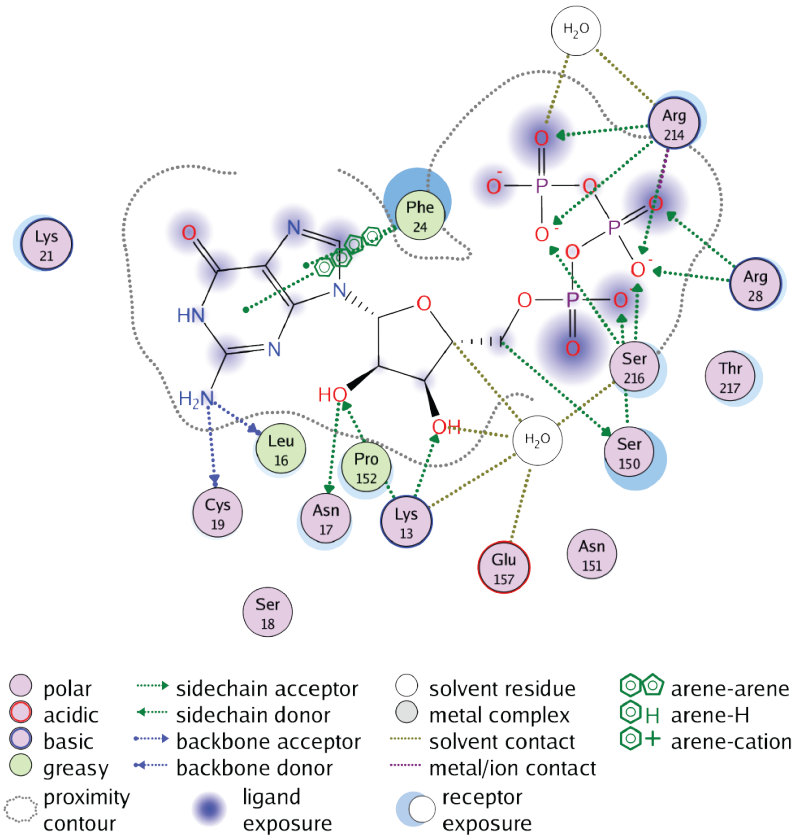


Figure 4.13 ***Ligand interaction plot of GTP in the GTP-binding site.***
A 2D representation of GTP in the GTP-binding site showing interactions with LVMT residues and ordered solvent. Figure generated with MOE 2009.10 (184).

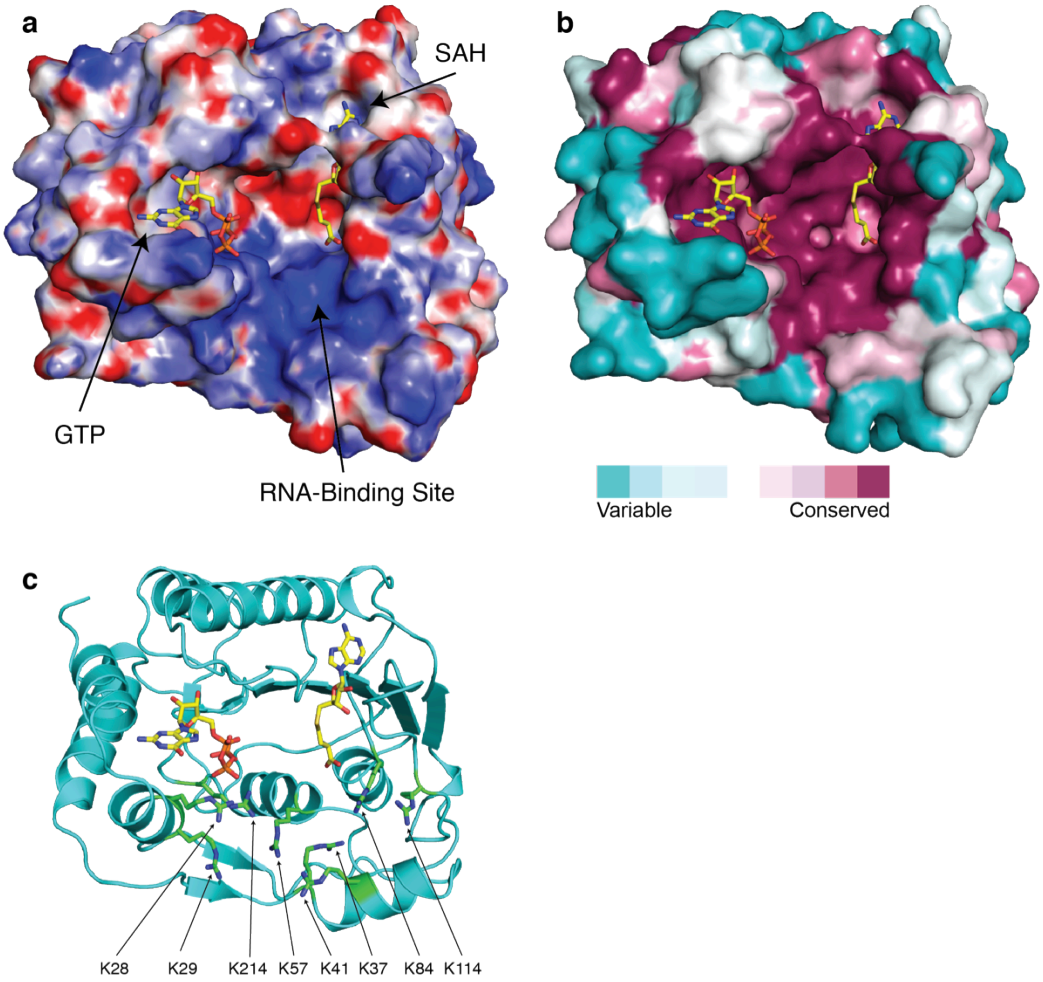


Figure 4.14 ***Putative LVMT RNA-binding site.***
(**a**) Electrostatic surface potential. The surface potential from +10 kT in red to -10 kT in blue is shown for the active site cleft. (**b**) Conserved active site cleft. The most conserved surface of LVMT is the active site cleft and the conservation extends to both SAM- and GTP-binding sites. (**c**) Invariant lysine residues. The LVMT is rendered as a green ribbon and putative RNA-binding site lysine residues rendered in stick form with cyan C atoms. SAH and GTP, rendered in stick form with yellow carbon atoms, are shown in their binding sites.

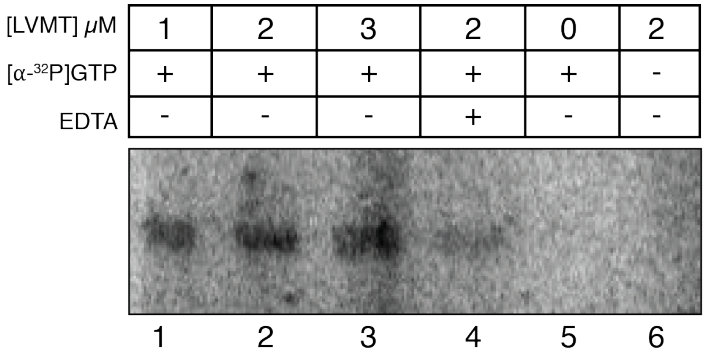


Figure 4.15 ***Formation of covalent LVMT-GMP adduct.***
Lane 1-3: LVMT was incubated with 1 μM, 2 μM and 3 μM of LVMT with 1 mM [α-^{32}P]GTP (PerkinElmer) in a buffer containing 20 mM Tris pH 7.8, 100 mM NaCl, 5 mM DTT, 5 mM $MgCl_2$ at 30°C for 1 hour. Lane 4: 10 mM EDTA was added to the reaction mixture before incubation. Lane 5-6: Negative controls with no protein or no [α-^{32}P]GTP. The reactions were quenched by addition of SDS-PAGE running buffer containing 1% SDS. Reaction samples were analyzed by electrophoresis with a 12% polyacrylamide gel containing 0.1% SDS. Radiolabeled proteins were visualized by autoradiography.

Discussion

This study investigated structural and functional differences between mosquito- and tick-borne flaviviruses MTase domains. We performed structural analysis of the tick-borne Langat virus MTase cap-binding and preliminary biochemical characterization of the LVMT guanyltransferase activity. High-resolution structures of LVMT bound to GTP and the GpppA CAP analogue provided insights in GTP binding and the correct orientation of RNA substrates for ribose 2′-O methy transfer.

Crystal structures of LVMT soaked with the GpppA cap analogue showed an unusual internal stacking interaction. The adenine base stacked with the GTP cap, resulting in an RNA conformation that in incompatable with guanine N7 or ribose 2′-O methylation (Fig. 4.12b) (178). However, co-crystallization of the cap analogues with the LVMT positioned the cap in a conformation that would allow the adenosine to be correctly oriented for 2′-O-methylation (Figs 4.12c). Poor electron density for adenosine in the active site cleft suggests mobility and a lack of specific interactions with the protein. The putative RNA-binding surface may stabilize the adenosine in the active site cleft by binding downstream nucleotides in the RNA substrate (Fig 4.14). With both the cap and downstream nucleotides bound to the MTase, the adenosine would be more restricted and susceptible to 2′-O methylation by SAM.

A structural explanation for the guanine N7 methyltransferase activity of flavivirus MTases has not been provided. Flavivirus MTases catalyze both guanine N7 and ribose 2′-O methylations during viral cap formation (65, 176). Crystal structures of GTP or cap analogues bound to flavivirus MTase show the GTP-binding site for ribose 2′-O methylation (Fig. 4.12) (66, 186). This binding is incompatible with N7 methylation, so an alternative GTP binding site is required for the MTase to methylate the guanine cap. Because all GTP- or cap-bound flavivirus MTase structures to date were crystal soaks, the MTases in the crystalline state may preclude access to an alternative GTP-binding site. We attempted to identify alternative GTP-binding sites by incubating LVMT with GTP or cap analogue prior to

crystallization. However, GTP and the cap analogues all bound to the same GTP-binding site seen in all MTase crystal structures (Fig. 4.12) (66, 176, 186).

Flavivirus guanine N7 methylation activity has been demonstrated in WNV, DENV and YFV MTases (67, 176). In those studies, N7 methylation only occurred with RNA substrates at least 71 nucleotides long (176). Other groups have been unable to demonstrate N7 methylation for flaviviruses with shorter RNA substrates or with cap analogues (65). Furthermore, viral RNA sequences with specific secondary structural elements in a 5′ stem-loop were required for efficient N7 methylation (187). Therefore, flaviviruses may use an alternative mechanism to methylate the N7 of the cap by utilizing the MTase domain and distinct RNA secondary-structure elements at the 5′ end of the genome.

In addition to guanine N7 and ribose 2'-O methylatransferase activity, flavivirus MTases have been shown to possess guanylyltransferase (GTase) activity (178). Recent biochemical studies of DENV, YFV, and WNV revealed that flavivirus MTase domains perform the actual capping of the RNA genome (178). The capping reaction requires the formation of a covalent enzyme-GMP adduct intermediate followed by transfer of GMP to an RNA substrate that has been processed to remove the γ phosphate (178). Our preliminary experiments using radiolabeled GTP showed that LVMT is also capable of forming the covalent intermediate and that the reaction is Mg^{2+} dependent (Fig. 4.15). Viral-encoded guanylyltransferases are well charecterized and structures are available for GTase-GMP adducts (188). However, flavivirus MTases lack sequence or structural homology to any known GTases or ATP-dependent DNA ligases, which share identical active site motifs with GTases (189). The molecular details of flavivirus guanylyltransferase activity are largely unknown, so further experiments are required to determine the efficiency of adduct formation, the effect of other flavivirus protein domains on the reaction efficiency and the guanylated residue in the MTase-GMP adduct.

Despite divergence in genomic sequence of the three flavivirus clades (179, 180), the active site cleft where SAM, GTP and RNA cap substrates bind has remained absolutely conserved in flavivirus MT domains (Figs 4.10 & 4.14b). The high degree

of conservation in the active site cleft illustrates the importance of a mature RNA cap for flavivirus viability.

Further examination of the interaction of flavivirus methyltransferases with their RNA substrates and with other viral proteins is required to better understand the capping process. Flavivirus NS5, which is both an RNA capping enzyme and an RNA-dependent RNA polymerase, is essential for the viability of flavivirus and is an attractive drug target, as inhibiting any of its functions should attenuate the virus.

CHAPTER 5
SUMMARY AND FUTURE DIRECTIONS

Genome packaging in Phleboviruses

Summary

RNP formation is essential for negative-sense RNA viruses during infections. RNPs protect the RNA genome from the antiviral response to ensure that the genome remain intact after infection. RNPs are also required for efficient transcription and genome replication by the viral RdRp. Crystal structures are available for helical RNPs from non-segmented RNA viruses, but the molecular details of N-RNA interactions in segmented virus RNP were unknown prior to this study.

In this study, we purified RVFV RNA-free N proteins using on-column protein refolding (Chapter 2). The RNA-free N allowed us to measure accurate RNA binding affinities using nucleic acids of varying sequence and lengths. We showed that N binds RNA with high affinity irrespective of sequence or length and binding affinities were affected by the nucleic acid content of the N preparation. These experiments also demonstrated that the high-affinity RNA-binding property of N was fully functional after refolding.

We showed that the monomeric RNA-free RVFV N has a novel fold unlike the N of other negative-sense RNA viruses (Chapter 2). It appeared to lack the deep, positively charged cleft that binds genomic RNA in other negative-sense RNA viruses. Also missing were extensions from the core that form inter-protein contacts to generate the helical RNP seen in other negative-sense viruses. These structural differences suggested that the structure of RVFV RNP might be significantly different from other negative-sense RNA viruses.

To investigate the structure of RVFV RNP, we used electron microscopy to image authentic RVFV RNP extracted from virus-infected cells (Chapter 2). Micrographs of viral RVFV RNP revealed a bead-on-a-string architecture different from the helical RNP observed by electron microscopy in other negative-sense RNA viruses. Reconstituted RNP (refolded N with a 680-nucleotide RNA strand) has an identical string-like appearance to the authentic RNP from virus-infected cells, showing that refolded N interacts with RNA similarly to N from virus-infected cells. Extensive

digestion of the string-like RNP with ribonuclease yielded small circular particles of similar size to the recombinant N-RNA complexes. However, these particles appeared heterogeneous, each containing between 4 and 7 protein subunits.

A clue to the structure of the small circular particles visualized by electron microscopy came from the N of Toscana virus (TOSV) (Chapter 3). We solved a crystal structure of TOSV N, which formed a ring-like hexamer similar to the multimers in the electron micrographs of RVFV RNP. As expected, TOSV N has the same overall protein fold as the RVFV N monomer, but a helical-arm structure detaches from the core of the protein and reaches around to interact with an adjacent N to form the hexamer ring. When separated from RNA, the TOSV N retains protein-protein contacts whereas the RVFV N does not. The helical-arm in TOSV N suggested that negative-sense RNA viruses use the same basic principle to assemble N proteins into RNPs. Protrusions from the core of N interact with neighboring N subunits to form an N polymer. N has an unlimited capacity for oligomerization, but the length of RNA determines RNP length.

We reported the structures of reconstituted RVFV N-RNA and N-DNA complexes (Chapter 3). These structures showed that N assembles into different circular oligomeric states depending on the length of the RNA. We discovered that the flexible helical arm accounts for the bead-on-a-string architecture of phlebovirus RNPs. While all other known structures of N from negative-sense RNA viruses restrict the movement of their arms in order to form symmetric helical structures, phleboviruses lack this rigidity, allowing for freedom of movement between N monomers, resulting in an asymmetric structure. The helical arm also enables N multimers to intrinsically form circles, which may be important for RdRp interaction.

Structures of reconstituted N-RNA showed that all RNA bases are sequestered into a deep RNA-binding slot (Chapter 3). While sequestered in the RNA-bind slot, RNA bases are inaccessible to other proteins including the viral RdRp. The binding slot is deep enough to accommodate purine or pyrimidine bases and correlates with the sequence independent genome packaging strategy of negative-sense RNA virus. The

"base-in" RNA arrangement is unique and implied that phlebovirus RdRp must strip N from the genome during transcription and replication. Together these data provide a detailed molecular picture of RNA packaging in phleboviruses that will help develop therapeutics to treat infections.

Future directions

The reconstituted RVFV N-RNA structures illustrated the molecular details of genome packaging, but several questions remain unanswered. We would like to investigate possible structural changes in RNA with purine bases in the RNA-binding slot. We are currently working towards a structure of N in complex with purine bases, which will reveal any differences in binding. The interaction between the RNP and the glycoproteins during viral assembly are also of interest. Mary Piper and Sonja Gerrard showed that the G_N cytoplasmic tail interacts with RNPs during viral assembly (31). We would like to characterize the interaction between the glycoproteins and the RNP through mutagenesis and binding experiments.

We would also like to investigate the interaction between N and the RdRp during replication and transcription. Currently, there are no crystal structures of negative-sense RNA virus RdRps. With the help of Clay Brown and Jim Delproposto at the High Throughput Protein Lab in the Life Sciences Institute, we would like to design plasmid for RdRp expression in insect cells. Purified RdRp will enable us to investigate RNP binding, replication and transcription in phleboviruses.

Finally, we would like to screen for compounds that will selectively bind in the RNA-binding slot of RVFV N and prevent RNA binding. The fluorescence polarization assay will be converted to a high-throughput screening assay for these experiments.

Genome capping in flaviviruses

Summary

We described the first structure of a tick-borne flavivirus methyltransferase domain (Chapter 4). The Langat virus methyltransferase domain is structurally homologous to other flavivirus MTases with known structure. Structures of GTP and cap-bound MTase domain showed the internal stacking observed in crystals of DENV and YFV MTases soaked with cap analogues. However, we discovered that cap analogues co-crystallized with the LVMT resulted in a cap orientation that is compatible with 2′-O methylation.

Preliminary experiments demonstrated the guanylyltransferase activity of LVMT, which is Mg^{2+} dependent (Chapter 4). These results demonstrated that mosquito- and tick-borne flavivirus MTases have identical functions.

Overall, our results showed that the active site cleft in flavivirus MTases is highly conserved despite divergence in the genomic sequence. The conservation of the active site is critical to virus viability because a mature cap is required for viability. Therefore, drugs designed to inhibit any of the MTase functions will treat infections by all flaviviruses.

Future directions

An MTase-RNA cap complex is not available with RNA bound in the active site cleft in a conformation compatible with methylation by SAM in the SAM-binding site. A structure of DENV MTase bound to capped RNA with eight nucleotides had the identical internal base stacking observed in the cap-soaked crystals (190). The first base stacks with the GTP cap and orients the seven other RNA nucleotides away from the SAM-binding site (190). Based on our experiments, co-crystallizing LVMT and capped RNA with more nucleotides may position the RNA substrate correctly for 2′-O methylation of the first base.

The molecular details of the Guanylyltransferase activity of flaviviruses are still unknown. The guanylyltransferase assay must be optimized to increase the guanyl

transfer efficiency. Biochemical studies showed that only ~5% of DENV, YFV or WNV MTase domains were guanylated using an identical guanylyltransferase assay (178). The low GTase efficiency indicated that flavivirus MTase are either inefficient GTases or require other viral or host proteins for optimal activity. Mass spectrometry experiments of LVMT incubated with GTP and Mg^{2+} were unable to identify guanylated residues. If optimization results in higher guanyl transfer, mass spectrometry experiments will be repeated to identify guanylated residue(s) in the MTase domain. Otherwise, the GTase assay with radiolabeled GTP will be used for future experiments.

REFERENCES

1. Weaver SC (2006) Evolutionary influences in arboviral disease. *Curr Top Microbiol Immunol* 299:285-314.

2. Drake JW & Holland JJ (1999) Mutation rates among RNA viruses. *Proc Natl Acad Sci U S A* 96:13910-13913.

3. Ciota AT & Kramer LD (2010) Insights into arbovirus evolution and adaptation from experimental studies. *Viruses* 2:2594-2617.

4. Gritsun TS, Nuttall PA, & Gould EA (2003) Tick-borne flaviviruses. *Adv Virus Res* 61:317-371.

5. Kuno G & Chang GJ (2005) Biological transmission of arboviruses: reexamination of and new insights into components, mechanisms, and unique traits as well as their evolutionary trends. *Clin Microbiol Rev* 18:608-637.

6. Marra PP, Griffing SM, & McLean RG (2003) West Nile virus and wildlife health. *Emerg Infect Dis* 9:898-899.

7. Weaver SC & Reisen WK (2010) Present and future arboviral threats. *Antiviral Res* 85:328-345.

8. Gould EA & Higgs S (2009) Impact of climate change and other factors on emerging arbovirus diseases. *Trans R Soc Trop Med Hyg* 103:109-121.

9. Ribeiro JM (1995) Blood-feeding arthropods: live syringes or invertebrate pharmacologists? *Infect Agents Dis* 4:143-152.

10. Barrett AD & Teuwen DE (2009) Yellow fever vaccine - how does it work and why do rare cases of serious adverse events take place? *Curr Opin Immunol* 21:308-313.

11. Heinz FX & Stiasny K (2012) Flaviviruses and flavivirus vaccines. *0264-410X* 30:4301-4306.

12. Stock NK, Boschetti N, Herzog C, Appelhans MS, & Niedrig M (2012) The phylogeny of yellow fever virus 17D vaccines. *0264-410X* 30:989-994.

13. Beaty BJ (2005) Control of arbovirus diseases: is the vector the weak link? *Arch Virol Suppl*:73-88.

14. Roberts DR, Laughlin LL, Hsheih P, & Legters LJ (1997) DDT, global strategies, and a malaria control crisis in South America. *Emerg Infect Dis* 3:295-302.

15. Daubney R, Hudson JR, & Garnham PC (1931) Enzootic hepatitis or rift valley fever. An undescribed virus disease of sheep cattle and man from east africa. *The Journal of Pathology and Bacteriology* 34:545-579.

16. El-Din Abdel-Wahab KS*, et al.* (1978) Rift Valley Fever virus infections in Egypt: pathological and virological findings in man. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 72:392-396.

17. Davies FG, Linthicum KJ, & James AD (1985) Rainfall and epizootic Rift Valley fever. *Bull World Health Organ* 63:941-943.

18. Philippe B*, et al.* (1989) Hemorrhagic forms of Rift Valley fever in Mauritania. *Bull Soc Pathol Exot Filiales* 82:611-619.

19. Arthur RR*, et al.* (1993) Recurrence of Rift Valley fever in Egypt. *Lancet* 342:1149-1150.

20. Fontenille D*, et al.* (1998) New vectors of Rift Valley fever in West Africa. *Emerg Infect Dis* 4:289-293.

21. Meegan JM (1979) The Rift Valley fever epizootic in Egypt 1977-78. 1. Description of the epizzotic and virological studies. *Trans R Soc Trop Med Hyg* 73:618-623.

22. Bouloy M & Flick R (2009) Reverse genetics technology for Rift Valley fever virus: current and future applications for the development of therapeutics and vaccines. *Antiviral Res* 84:101-118.

23. Meegan JM, Hoogstraal H, & Moussa MI (1979) An epizootic of Rift Valley fever in Egypt in 1977. *Vet Rec* 105:124-125.

24. Billecocq A, Vialat P, & Bouloy M (1996) Persistent infection of mammalian cells by Rift Valley fever virus. *J Gen Virol* 77 (Pt 12):3053-3062.

25. Zeller HG, Fontenille D, Traore-Lamizana M, Thiongane Y, & Digoutte JP (1997) Enzootic activity of Rift Valley fever virus in Senegal. *Am J Trop Med Hyg* 56:265-272.

26. Freed I (1951) Rift valley fever in man, complicated by retinal changes and loss of vision. *S Afr Med J* 25:930-932.

27. Madani TA*, et al.* (2003) Rift Valley fever epidemic in Saudi Arabia: epidemiological, clinical, and laboratory characteristics. *Clin Infect Dis* 37:1084-1092.

28. Fields BN, Knipe DM, Howley PM, & Griffin DE (2001). *Fields Viroogy*, Lippincott, Williams & Wilkins, Philadelphia, pp 3087, 3072.

29. Bouloy M & Weber F (2010) Molecular biology of rift valley Fever virus. *Open Virol J* 4:8-14.

30. Guu TS, Zheng W, & Tao YJ (2012) Bunyavirus: structure and replication. *Adv Exp Med Biol* 726:245-266.

31. Piper ME, Sorenson DR, & Gerrard SR (2011) Efficient Cellular Release of Rift Valley Fever Virus Requires Genomic RNA. *PLoS ONE* 6:e18070.

32. Overby AK, Pettersson RF, & Neve EP (2007) The glycoprotein cytoplasmic tail of Uukuniemi virus (Bunyaviridae) interacts with ribonucleoproteins and is critical for genome packaging. *J Virol* 81:3198-3205.

33. Ruigrok RW, Crepin T, & Kolakofsky D (2011) Nucleoproteins and nucleocapsids of negative-strand RNA viruses. *Curr Opin Microbiol* 14:504-510.

34. Westaway EG, *et al.* (1985) Flaviviridae. *Intervirology* 24:183-192.

35. Lindenbach D & Rice CM (2001). *Fields Virology*, eds Lamb RA, Martin MA, Roizman B, & Strauss SE Lippincott-Raven Publishers, Philadelphia, pp 991-1041.

36. CDC (2007) Dengue Fever Fact Sheet. (Center for Disease Control and Prevention).

37. Gubler DJ (2002) The global emergence/resurgence of arboviral diseases as public health problems. *Arch Med Res* 33:330-342.

38. Gubler DJ (2002) Epidemic dengue/dengue hemorrhagic fever as a public health, social and economic problem in the 21st century. *Trends Microbiol* 10:100-103.

39. Rice CM, *et al.* (1985) Nucleotide sequence of yellow fever virus: implications for flavivirus gene expression and evolution. *Science* 229:726-733.

40. Wengler G, Nowak T, & Castle E (1990) Description of a procedure which allows isolation of viral nonstructural proteins from BHK vertebrate cells infected with the West Nile flavivirus in a state which allows their direct chemical characterization. *Virology* 177:795-801.

41. Cauchi MR, Henchal EA, & Wright PJ (1991) The sensitivity of cell-associated dengue virus proteins to trypsin and the detection of trypsin-resistant fragments of the nonstructural glycoprotein NS1. *Virology* 180:659-667.

42. Chambers TJ, Hahn CS, Galler R, & Rice CM (1990) Flavivirus genome organization, expression, and replication. *Annu Rev Microbiol* 44:649-688.

43. Chambers TJ, Grakoui A, & Rice CM (1991) Processing of the yellow fever virus nonstructural polyprotein: a catalytically active NS3 proteinase domain and NS2B are required for cleavages at dibasic sites. *J Virol* 65:6042-6050.

44. Ryan MD, Monaghan S, & Flint M (1998) Virus-encoded proteinases of the Flaviviridae. *J Gen Virol* 79 (Pt 5):947-959.

45. Kong KF, Wang X, Anderson JF, Fikrig E, & Montgomery RR (2008) West nile virus attenuates activation of primary human macrophages. *Viral Immunol* 21:78-82.

46. Munoz-Jordan JL*, et al.* (2005) Inhibition of alpha/beta interferon signaling by the NS4B protein of flaviviruses. *J Virol* 79:8004-8013.

47. Munoz-Jordan JL, Sanchez-Burgos GG, Laurent-Rolle M, & Garcia-Sastre A (2003) Inhibition of interferon signaling by dengue virus. *Proc Natl Acad Sci U S A* 100:14333-14338.

48. Li H, Clum S, You S, Ebner KE, & Padmanabhan R (1999) The serine protease and RNA-stimulated nucleoside triphosphatase and RNA helicase functional domains of dengue virus type 2 NS3 converge within a region of 20 amino acids. *J Virol* 73:3108-3116.

49. Preugschat F, Yao CW, & Strauss JH (1990) In vitro processing of dengue virus type 2 nonstructural proteins NS2A, NS2B, and NS3. *J Virol* 64:4364-4374.

50. Kadare G & Haenni AL (1997) Virus-encoded RNA helicases. *J Virol* 71:2583-2590.

51. Wengler G (1993) The NS 3 nonstructural protein of flaviviruses contains an RNA triphosphatase activity. *Virology* 197:265-273.

52. Gorbalenya AE & Koonin EV (1989) Viral proteins containing the purine NTP-binding sequence pattern. *Nucleic Acids Res* 17:8413-8440.

53. de la Cruz J, Kressler D, & Linder P (1999) Unwinding RNA in Saccharomyces cerevisiae: DEAD-box proteins and related families. *Trends Biochem Sci* 24:192-198.

54. Walker JE, Saraste M, Runswick MJ, & Gay NJ (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J* 1:945-951.

55. Cho HS, *et al.* (1998) Crystal structure of RNA helicase from genotype 1b hepatitis C virus. A feasible mechanism of unwinding duplex RNA. *J Biol Chem* 273:15045-15052.

56. Tseng-Rogenski SS & Chang TH (2004) RNA unwinding assay for DExD/H-box RNA helicases. *Methods Mol Biol* 257:93-102.

57. Suzich JA, *et al.* (1993) Hepatitis C virus NS3 protein polynucleotide-stimulated nucleoside triphosphatase and comparison with the related pestivirus and flavivirus enzymes. *J Virol* 67:6152-6158.

58. Warrener P, Tamura JK, & Collett MS (1993) RNA-stimulated NTPase activity associated with yellow fever virus NS3 protein expressed in bacteria. *J Virol* 67:989-996.

59. Wengler G (1991) The carboxy-terminal part of the NS 3 protein of the West Nile flavivirus can be isolated as a soluble protein after proteolytic cleavage and represents an RNA-stimulated NTPase. *Virology* 184:707-715.

60. Mancini EJ, *et al.* (2007) Structure of the Murray Valley encephalitis virus RNA helicase at 1.9 Angstrom resolution. *Protein Sci* 16:2294-2300.

61. Wu J, Bera AK, Kuhn RJ, & Smith JL (2005) Structure of the Flavivirus helicase: implications for catalytic activity, protein interactions, and proteolytic processing. *J Virol* 79:10268-10277.

62. Xu T, *et al.* (2005) Structure of the Dengue virus helicase/nucleoside triphosphatase catalytic domain at a resolution of 2.4 A. *J Virol* 79:10278-10288.

63. Koonin EV (1991) The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *J Gen Virol* 72 (Pt 9):2197-2206.

64. Koonin EV (1993) Computer-assisted identification of a putative methyltransferase domain in NS5 protein of flaviviruses and lambda 2 protein of reovirus. *J Gen Virol* 74 (Pt 4):733-740.

65. 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.

66. Egloff MP, *et al.* (2007) Structural and functional analysis of methylation and 5'-RNA sequence requirements of short capped RNAs by the methyltransferase domain of dengue virus NS5. *J Mol Biol* 372:723-736.

67. Ray D, *et al.* (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.

68. Chu PW & Westaway EG (1987) Characterization of Kunjin virus RNA-dependent RNA polymerase: reinitiation of synthesis in vitro. *Virology* 157:330-337.

69. Bartholomeusz AI & Wright PJ (1993) Synthesis of dengue virus RNA in vitro: initiation and the involvement of proteins NS3 and NS5. *Arch Virol* 128:111-121.

70. Tan BH, *et al.* (1996) Recombinant dengue type 1 virus NS5 protein expressed in Escherichia coli exhibits RNA-dependent RNA polymerase activity. *Virology* 216:317-325.

71. Ackermann M & Padmanabhan R (2001) De novo synthesis of RNA by the dengue virus RNA-dependent RNA polymerase exhibits temperature dependence at the initiation but not elongation phase. *J Biol Chem* 276:39926-39937.

72. Guyatt KJ, Westaway EG, & Khromykh AA (2001) Expression and purification of enzymatically active recombinant RNA-dependent RNA polymerase (NS5) of the flavivirus Kunjin. *J Virol Methods* 92:37-44.

73. Steffens S, Thiel HJ, & Behrens SE (1999) The RNA-dependent RNA polymerases of different members of the family Flaviviridae exhibit similar properties in vitro. *J Gen Virol* 80 (Pt 10):2583-2590.

74. Khromykh AA, Kenney MT, & Westaway EG (1998) trans-Complementation of flavivirus RNA polymerase gene NS5 by using Kunjin virus replicon-expressing BHK cells. *J Virol* 72:7270-7279.

75. Buckley A, Gaidamovich S, Turchinskaya A, & Gould EA (1992) Monoclonal antibodies identify the NS5 yellow fever virus non-structural protein in the nuclei of infected cells. *J Gen Virol* 73 (Pt 5):1125-1130.

76. Kapoor M, *et al.* (1995) Association between NS3 and NS5 proteins of dengue virus type 2 in the putative RNA replicase is linked to differential phosphorylation of NS5. *J Biol Chem* 270:19100-19106.

77. Johansson M, Brooks AJ, Jans DA, & Vasudevan SG (2001) A small region of the dengue virus-encoded RNA-dependent RNA polymerase, NS5, confers interaction with both the nuclear transport receptor importin-beta and the viral helicase, NS3. *J Gen Virol* 82:735-745.

78. Uchil PD, Kumar AV, & Satchidanandam V (2006) Nuclear localization of flavivirus RNA synthesis in infected cells. *J Virol* 80:5451-5464.

79. Cui T*, et al.* (1998) Recombinant dengue virus type 1 NS3 protein exhibits specific viral RNA binding and NTPase activity regulated by the NS5 protein. *Virology* 246:409-417.

80. Varani G (1997) A cap for all occasions. *Structure* 5:855-858.

81. Flint SJ, Enquist LW, Racaniello VR, & Skalka AM (2004). *Principles of virology : molecular biology, pathogenesis, and control of animal viruses*, ASM Press, Washington, D.C., pp 83-125.

82. Nichol ST (2001). *Fields Virology*, eds Fields BN, Knipe DM, Howley PM, & Griffin DE Lippincott, Williams & Wilkins, Philadelphia, PA, pp 1603-1633.

83. Schmaljohn CS & Hooper JW (2001). *Fields Virology*, eds Fields BN, Knipe DM, Howley PM, & Griffin DE Lippincott, Williams & Wilkins, Philadelphia, PA, pp 1581-1602.

84. Schmaljohn CS & Nichol ST (2007). *Fields Virology*, eds Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MA, Roizman B, & Straus SE Lippincott, Williams & Wilkins, Philadelphia, PA, pp 1741-1789.

85. Mohl BP & Barr JN (2009) Investigating the specificity and stoichiometry of RNA binding by the nucleocapsid protein of Bunyamwera virus. *RNA* 15:391-399.

86. Gott P, Stohwasser R, Schnitzler P, Darai G, & Bautz EK (1993) RNA binding of recombinant nucleocapsid proteins of hantaviruses. *Virology* 194:332-337.

87. Osborne JC & Elliott RM (2000) RNA binding properties of bunyamwera virus nucleocapsid protein and selective binding to an element in the 5' terminus of the negative-sense S segment. *J Virol* 74:9946-9952.

88. Mir MA, Brown B, Hjelle B, Duran WA, & Panganiban AT (2006) Hantavirus N protein exhibits genus-specific recognition of the viral RNA panhandle. *J Virol* 80:11283-11292.

89. Ogg MM & Patterson JL (2007) RNA binding domain of Jamestown Canyon virus S segment RNAs. *J Virol* 81:13754-13760.

90. Severson W, Partin L, Schmaljohn CS, & Jonsson CB (1999) Characterization of the Hantaan nucleocapsid protein-ribonucleic acid interaction. *J Biol Chem* 274:33732-33739.

91. Falk BW & Tsai JH (1998) Biology and molecular biology of viruses in the genus Tenuivirus. *Annu Rev Phytopathol* 36:139-163.

92. Ye Q, Krug RM, & Tao YJ (2006) The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA. *Nature* 444:1078-1082.

93. Albertini AA, *et al.* (2006) Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science* 313:360-363.

94. Tawar RG, *et al.* (2009) Crystal Structure of a Nucleocapsid-Like Nucleoprotein-RNA Complex of Respiratory Syncytial Virus. *Science* 326:1279-1283.

95. Green TJ, Zhang X, Wertz GW, & Luo M (2006) Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. *Science* 313:357-360.

96. Rudolph MG, *et al.* (2003) Crystal structure of the borna disease virus nucleoprotein. *Structure* 11:1219-1226.

97. Coloma R, *et al.* (2009) The structure of a biologically active influenza virus ribonucleoprotein complex. *PLoS Pathog* 5:e1000491.

98. von Bonsdorff CH, Saikku P, & Oker-Blom N (1969) The inner structure of Uukuniemi and two Bunyamwera supergroup arboviruses. *Virology* 39:342-344.

99. Pettersson RF & von Bonsdorff CH (1975) Ribonucleoproteins of Uukuniemi virus are circular. *J Virol* 15:386-392.

100. Schoehn G, *et al.* (2004) The 12 A structure of trypsin-treated measles virus N-RNA. *J Mol Biol* 339:301-312.

101. Bhella D, Ralph A, Murphy LB, & Yeo RP (2002) Significant differences in nucleocapsid morphology within the Paramyxoviridae. *J Gen Virol* 83:1831-1839.

102. Ge P, *et al.* (2010) Cryo-EM model of the bullet-shaped vesicular stomatitis virus. *Science* 327:689-693.

103. Ruigrok RW & Baudin F (1995) Structure of influenza virus ribonucleoprotein particles. II. Purified RNA-free influenza virus ribonucleoprotein forms structures that are indistinguishable from the intact influenza virus ribonucleoprotein particles. *J Gen Virol* 76 (Pt 4):1009-1014.

104. Gerrard SR, Bird BH, Albarino CG, & Nichol ST (2007) The NSm proteins of Rift Valley fever virus are dispensable for maturation, replication and infection. *Virology* 359:459-465.

105. Gerrard SR & Nichol ST (2002) Characterization of the Golgi retention motif of Rift Valley fever virus G_N glycoprotein. *J Virol* 76:12200-12210.

106. Rubach JK*, et al.* (2009) Characterization of purified Sindbis virus nsP4 RNA-dependent RNA polymerase activity in vitro. *Virology* 384:201-208.

107. Marenchino M, Armbruster DW, & Hennig M (2009) Rapid and efficient purification of RNA-binding proteins: application to HIV-1 Rev. *Protein Expr Purif* 63:112-119.

108. Reverter D & Lima CD (2004) A basis for SUMO protease specificity provided by analysis of human Senp2 and a Senp2-SUMO complex. *Structure* 12:1519-1531.

109. Otwinowski Z & Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Meth Enzymol* 276:307-326.

110. Battye TG, Kontogiannis L, Johnson O, Powell HR, & Leslie AG (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* 67:271-281.

111. Leslie AG (2006) The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr* 62:48-57.

112. Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* 62:72-82.

113. Terwilliger TC & Berendzen J (1999) Automated MAD and MIR structure solution. *Acta Crystallogr D Biol Crystallogr* 55:849-861.

114. Terwilliger TC (2003) Automated main-chain model building by template matching and iterative fragment extension. *Acta Crystallogr D Biol Crystallogr* 59:38-44.

115. Adams PD*, et al.* (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* 58:1948-1954.

116. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126-2132.

117. Murshudov GN, Vagin AA, & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53:240-255.

118. Collaborative Computational Project N (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50:760-763.

119. Davis IW*, et al.* (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35:W375-383.

120. Delano WL (2007) The PyMOL Molecular Graphics System (DeLano Scientific LLC, USA).

121. Thompson JD, Gibson TJ, & Higgins DG (2002) Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics* Chapter 2:Unit 2 3.

122. 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.

123. Gouet P, Robert X, & Courcelle E (2003) ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res* 31:3320-3323.

124. Rai BK & Fiser A (2006) Multiple mapping method: a novel approach to the sequence-to-structure alignment problem in comparative protein structure modeling. *Proteins* 63:644-661.

125. Landau M*, et al.* (2005) ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res* 33:W299-302.

126. Piper ME & Gerrard SR (2010) A novel system for identification of inhibitors of rift valley Fever virus replication. *Viruses* 2:731-747.

127. Lamanna AC & Karbstein K (2009) Nob1 binds the single-stranded cleavage site D at the 3'-end of 18S rRNA with its PIN domain. *Proc Natl Acad Sci U S A* 106:14259-14264.

128. Ohi M, Li Y, Cheng Y, & Walz T (2004) Negative Staining and Image Classification - Powerful Tools in Modern Electron Microscopy. *Biol Proced Online* 6:23-34.

129. Ludtke SJ, Baldwin PR, & Chiu W (1999) EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* 128:82-97.

130. Frank J*, et al.* (1996) SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J Struct Biol* 116:190-199.

131. Le May N, Gauliard N, Billecocq A, & Bouloy M (2005) The N terminus of Rift Valley fever virus nucleoprotein is essential for dimerization. *J Virol* 79:11974-11980.

132. Holm L & Sander C (1995) Dali: a network tool for protein structure comparison. *Trends Biochem Sci* 20:478-480.

133. Gibrat JF, Madej T, & Bryant SH (1996) Surprising similarities in structure comparison. *Curr Opin Struct Biol* 6:377-385.

134. Raju R & Kolakofsky D (1989) The ends of La Crosse virus genome and antigenome RNAs within nucleocapsids are base paired. *J Virol* 63:122-128.

135. Pettersson RF & Melin L (1996). *The Bunyaviridae*, ed Elliott RMPlenum, New York, pp 159-188.

136. Overby AK, Popov VL, Pettersson RF, & Neve EP (2007) The cytoplasmic tails of Uukuniemi Virus (Bunyaviridae) G_N and G_C glycoproteins are important for intracellular targeting and the budding of virus-like particles. *J Virol* 81:11381-11391.

137. Borucki MK, Chandler LJ, Parker BM, Blair CD, & Beaty BJ (1999) Bunyavirus superinfection and segment reassortment in transovarially infected mosquitoes. *J Gen Virol* 80:3173-3179.

138. Chandler LJ*, et al.* (1991) Reassortment of La Crosse and Tahyna bunyaviruses in Aedes triseriatus mosquitoes. *Virus Res* 20:181-191.

139. Beaty BJ, Sundin DR, Chandler LJ, & Bishop DH (1985) Evolution of bunyaviruses by genome reassortment in dually infected mosquitoes (Aedes triseriatus). *Science* 230:548-550.

140. Briese T, Kapoor V, & Lipkin WI (2007) Natural M-segment reassortment in Potosi and Main Drain viruses: implications for the evolution of orthobunyaviruses. *Arch Virol* 152:2237-2247.

141. Bowen MD*, et al.* (2001) A reassortant bunyavirus isolated from acute hemorrhagic fever cases in Kenya and Somalia. *Virology* 291:185-190.

142. Pepin M, Bouloy M, Bird BH, Kemp A, & Paweska J (2010) Rift Valley fever virus(Bunyaviridae: Phlebovirus): an update on pathogenesis, molecular epidemiology, vectors, diagnostics and prevention. *Vet Res* 41:61.

143. Valassina M, Cusi MG, & Valensin PE (2003) A Mediterranean arbovirus: the Toscana virus. *J Neurovirol* 9:577-583.

144. Honda A, Ueda K, Nagata K, & Ishihama A (1988) RNA polymerase of influenza virus: role of NP in RNA chain elongation. *J Biochem* 104:1021-1026.

145. 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.

146. Ferron F, *et al.* (2011) The hexamer structure of Rift Valley fever virus nucleoprotein suggests a mechanism for its assembly into ribonucleoprotein complexes. *PLoS Pathog* 7:e1002030.

147. Weeks SD, Drinker M, & Loll PJ (2007) Ligation independent cloning vectors for expression of SUMO fusions. *Protein Expr Purif* 53:40-50.

148. Motulsky HJ & Neubig RR (2010) Analyzing binding data. *Current Protocols in Neuroscience* Unit 7.5.

149. McCoy AJ, *et al.* (2007) Phaser crystallographic software. *J Appl Crystallogr* 40:658-674.

150. Keating KS & Pyle AM (2010) Semiautomated model building for RNA crystallography using a directed rotameric approach. *Proc Natl Acad Sci U S A* 107:8177-8182.

151. Bricogne G, *et al.* (2010) BUSTER (Global Phasing Ltd, Cambridge, United Kingdom), 2.9.

152. Painter J & Merritt EA (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr* 62:439-450.

153. Hastie KM, *et al.* (2011) Crystal structure of the Lassa virus nucleoprotein-RNA complex reveals a gating mechanism for RNA binding. *Proc Natl Acad Sci U S A* 108:19365-19370.

154. Hewlett MJ, Pettersson RF, & Baltimore D (1977) Circular forms of Uukuniemi virion RNA: an electron microscopic study. *J Virol* 21:1085-1093.

155. Young PR & Howard CR (1983) Fine structure analysis of Pichinde virus nucleocapsids. *J Gen Virol* 64 (Pt 4):833-842.

156. Hastie KM, Kimberlin CR, Zandonatti MA, MacRae IJ, & Saphire EO (2011) Structure of the Lassa virus nucleoprotein reveals a dsRNA-specific 3' to 5' exonuclease activity essential for immune suppression. *Proc Natl Acad Sci U S A* 108:2396-2401.

157. Qi X, *et al.* (2010) Cap binding and immune evasion revealed by Lassa nucleoprotein structure. *Nature* 468:779-783.

158. Freiberg AN, Sherman MB, Morais MC, Holbrook MR, & Watowich SJ (2008) Three-dimensional organization of Rift Valley fever virus revealed by cryoelectron tomography. *J Virol* 82:10341-10348.

159. Klumpp K, Ruigrok RW, & Baudin F (1997) Roles of the influenza virus polymerase and nucleoprotein in forming a functional RNP structure. *EMBO J* 16:1248-1257.

160. Noda T & Kawaoka Y (2010) Structure of influenza virus ribonucleoprotein complexes and their packaging into virions. *Rev Med Virol* 20:380-391.

161. Voss NR & Gerstein M (2005) Calculation of standard atomic volumes for RNA and comparison with proteins: RNA is packed more tightly. *J Mol Biol* 346:477-492.

162. Gauliard N, Billecocq A, Flick R, & Bouloy M (2006) Rift Valley fever virus noncoding regions of L, M and S segments regulate RNA synthesis. *Virology* 351:170-179.

163. Obijeski JF, Bishop DH, Palmer EL, & Murphy FA (1976) Segmented genome and nucleocapsid of La Crosse virus. *J Virol* 20:664-675.

164. Kohl A, Dunn EF, Lowen AC, & Elliott RM (2004) Complementarity, sequence and structural elements within the 3' and 5' non-coding regions of the Bunyamwera orthobunyavirus S segment determine promoter strength. *J Gen Virol* 85:3269-3278.

165. Flick K*, et al.* (2004) Functional analysis of the noncoding regions of the Uukuniemi virus (Bunyaviridae) RNA segments. *J Virol* 78:11726-11738.

166. Kuhn RJ (2006). *In Fundamentals of Molecular Virology*, ed Acheson NH John Wiley & Sons, Inc, pp 181-190.

167. Weaver SC (2006) Evolutionary influences in arboviral disease. *Current topics in microbiology and immunology* 299:285-314.

168. Brinton MA (2002) The molecular biology of West Nile Virus: a new invader of the western hemisphere. *Annu Rev Microbiol* 56:371-402.

169. Whitehorn J & Farrar J (2010) Dengue. *Br Med Bull* 95:161-173.

170. Guzman MG*, et al.* (2010) Dengue: a continuing global threat. *Nat Rev Microbiol* 8:S7-16.

171. Ranjit S & Kissoon N (2011) Dengue hemorrhagic fever and shock syndromes. *Pediatr Crit Care Med* 12:90-100.

172. Smith CE (1956) A virus resembling Russian spring-summer encephalitis virus from an ixodid tick in Malaya. *Nature* 178:581-582.

173. Rumyantsev AA, Murphy BR, & Pletnev AG (2006) A tick-borne Langat virus mutant that is temperature sensitive and host range restricted in neuroblastoma cells and lacks neuroinvasiveness for immunodeficient mice. *J Virol* 80:1427-1439.

174. Thind S (1981) Potential use of attenuated langat E5 virus as a live vaccine -- long term protection against Russian spring-summer encephalitis virus in mice. *J Hyg Epidemiol Microbiol Immunol* 25:155-162.

175. Chambers TJ, McCourt DW, & Rice CM (1990) Production of yellow fever virus proteins in infected cells: identification of discrete polyprotein species and analysis of cleavage kinetics using region-specific polyclonal antisera. *Virology* 177:159-174.

176. Zhou Y, *et al.* (2007) Structure and function of flavivirus NS5 methyltransferase. *J Virol* 81:3891-3903.

177. Benarroch D, *et al.* (2004) A structural basis for the inhibition of the NS5 dengue virus mRNA 2'-O-methyltransferase domain by ribavirin 5'-triphosphate. *J Biol Chem* 279:35638-35643.

178. Issur M, *et al.* (2009) The flavivirus NS5 protein is a true RNA guanylyltransferase that catalyzes a two-step reaction to form the RNA cap structure. *RNA* 15:2340-2350.

179. Gaunt MW, *et al.* (2001) Phylogenetic relationships of flaviviruses correlate with their epidemiology, disease association and biogeography. *J Gen Virol* 82:1867-1876.

180. Kuno G, Chang GJ, Tsuchiya KR, Karabatsos N, & Cropp CB (1998) Phylogeny of the genus Flavivirus. *J Virol* 72:73-83.

181. Ecker M, Allison SL, Meixner T, & Heinz FX (1999) Sequence analysis and genetic classification of tick-borne encephalitis viruses from Europe and Asia. *J Gen Virol* 80 (Pt 1):179-185.

182. Kabsch W (2010) Xds. *Acta Crystallographica Section D Biological Crystallography* 66:125-132.

183. Kabsch W (2010) Integration, scaling, space-group assignment and post-refinement. *Acta Crystallographica Section D Biological Crystallography* 66:133-144.

184. Group CC (2009) MOE (Molecular Operating Environment)Montreal, Canada), 2009.10.

185. Cheng X & Blumenthal R (1999) *S-adenosylmethionine-dependent methyltransferases: structures and functions*.

186. Geiss BJ*, et al.* (2009) Analysis of flavivirus NS5 methyltransferase cap binding. *J Mol Biol* 385:1643-1654.

187. Dong H*, et al.* (2007) Distinct RNA elements confer specificity to flavivirus RNA cap methylation events. *J Virol* 81:4412-4421.

188. Bisaillon M & Lemay G (1997) Viral and cellular enzymes involved in synthesis of mRNA cap structure. *Virology* 236:1-7.

189. Shuman S, Liu Y, & Schwer B (1994) Covalent catalysis in nucleotidyl transfer reactions: essential motifs in Saccharomyces cerevisiae RNA capping enzyme are conserved in Schizosaccharomyces pombe and viral capping enzymes and among polynucleotide ligases. *Proc Natl Acad Sci U S A* 91:12046-12050.

190. Yap LJ*, et al.* (2010) Crystal structure of the dengue virus methyltransferase bound to a 5'-capped octameric RNA. *PLoS One* 5.