Dengue Virus Non-structural Protein Complexes in Genome Replication: Structural and Functional Characterization

By

KUNDHARAPU SATYAMURTHY

LIFE11201504006

National Institute of Science Education and Research (NISER)
Bhubaneswar

A thesis submitted to the
Board of Studies in Life Sciences

In partial fulfilment of requirements

for the Degree of

DOCTOR OF PHILOSOPHY

Of

HOMI BHABHA NATIONAL INSTITUTE

January, 2023

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfilment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the Library to be made available to borrowers under rules of the HBNI.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

Kundharapu Satyamurthy

DECLARATION

I, hereby declare that the investigation presented in the thesis has been carried out by me. The work is original and has not been submitted earlier as a whole or in part for a degree / diploma at this or any other Institution / University.

Kundharapu Satyamurthy

List of publications arising from the thesis

Peer-reviewed Scientific Journals

1. Kundharapu S, Chowdary TK. Dengue Virus NS4b N-Terminus Disordered Region Interacts with NS3 Helicase C-Terminal Subdomain to Enhance Helicase Activity. Viruses. 2022 Aug 3;14(8):1712. doi: 10.3390/v14081712. PMID: 36016333; PMCID: PMC9412862. (Published)
2. Role of NS2b cofactor in NS3 protease activity of Dengue virus - a polyacidic sequence stretch in the NS2b as a key element in substrate binding and its translocation to the enzyme active site. Kundharapu S, Dora AS, Kapoor R and Chowdary TK (In preparation)

Chapters in books and lectures notes

None

Conferences

1. **Satymurthy Kundharapu**, Tirumala Kumar Chowdary. Dengue non-structural protein interaction in replication complex. 14th international conference on vectors and vectorborne diseases; 2019 January; Bhubaneshwar, Odisha.
2. **Satyamurthy Kundharapu**, Tirumala Kumar Chowdary. Structure, function characterization of dengue virus replication complex protein NS4B and its interaction with NS3 helicase. VIROCON 2020 International conference on the theme 'Evolution of Viruses and Viral Diseases'; 2020 February, New Delhi.
3. **Satyamurthy Kundharapu,** Armul Simanchal Dora, Tirumala Kumar Chowdary**.** Interaction of Dengue virus non-structural protein 3 protease domain with its cofactor, NS2b, and mechanism of substrate specificity. Serine proteases in pericellular proteolysis and signaling**,** 2021 October, Virtual, by ASBMB.

Others

None

Kundharapu Satyamurthy

ACKNOWLEDGEMENTS

First, I would like to extend my sincere gratitude to my mentor, Dr. Tirumala Kumar Chowdary for the valuable guidance and advice he has given me throughout my PhD and for the utmost freedom he has offered to me during entire course of the study, and for being always there to troubleshoot the problems, professional or otherwise.

I express my gratitude to our Director for providing the infrastructure and working facility for research at NISER. I am also thankful to my doctoral committee members: Prof. Chandan Goswami, Prof. Palok Aich, Dr. Rudresh Acharya, and Dr. Nagendra Kumar Sharma for their valuable inputs and suggestions during progress monitoring meetings. Further, I thank PGCS convener and Chairperson of school of biological sciences, Dean academics and all other faculties and staff members of NISER for their timely help during my Ph.D.

I would like to thank Dr. Easwaran Sreekumar (RGCB, Thiruvananthapuram) for providing us the plasmids of Dengue virus nonstructural proteins, and Dr. Ramanujam Srinivasan(NISER, Bhubaneshwar) and his lab members. For BACTH assay plasmids and protocols which I have used in my research work. I am really thankful to DAE-NISER for financial assistance through JRF and SRF fellowships during my Ph.D. I thank NISER and my mentor for providing financial assistance during the academic extension period.

I can never forget the support that I received from my lab members during the tenure of my Ph.D. All the members of our lab: Dr. Bibekananda sahoo, Dr. Anupriya goutham, Mr, Debashish panda, Mr. Deepak Ranjan Behra, Mr. Gudigamolla. Naresh Kumar, Mr Subham Kumar Tripathy, Mr. Sapuru Vinay Kumar, Ms. Preethi chanda Behra, Mr. Armul Simanchal Dora, Mr. Sourav Agarwal, Mr. Dheemanth Redyy, Ms. Hradini, Ms. Rupali Kapoor, Ms. Datti Ragasahithi, and Ms. Birudula Veena were really helpful and cordial.

I am very much thankful to my parents, brother and my wife for persistent moral support towards my success during Ph.D.

I thank ‘God’, above all, for giving me the strength to achieve this difficult milestone.

CONTENTS

Summary

The most important aspect of a successful viral infection is replication of its genome. Upon successful entry into a susceptible cell, viruses induce organization of specialized compartments in a cell and replicate their genetic material in those compartments. Majority of the newly synthesized viral proteins in a cell are recruited to these 'Replication Compartments/Organelles'(RO) and show very dynamic interactions amongst themselves. RNA viruses, such as Dengue virus (DENV), induce RO as invaginations into ER membranes. Dengue virus polymerase – an RNA dependent RNA polymerase – and a RNA helicase, named non-structural-5 (NS5) and NS3, respectively, are key components in the viral genome replication. NS3 interacts with several other viral proteins, NS2B, NS4A, NS4B, and these interactions are critical for the genome replication. However, the interaction interfaces and mechanism of the protein (and protein-RNA complexes) are not understood.

The primary objective of my thesis is to characterize the interactions amongst the NS proteins, and NS3 protein with different structural elements of the viral RNA to explain the mechanism of the protein-RNA complexes in replication. NS3 is a multi-domain, multi-function protein. The amino-terminal one-third of the protein has a canonical chymotrypsin fold and function as a serine protease along with another essential protein co-factor, NS2b. The C-terminus two-thirds of the protein functions as a RNA helicase with canonical super family 2 (SF2) RNA helicase motifs. We investigated the interactions between NS3 and other nonstructural proteins. We focused on three major interactions of NS3 with other nonstructural proteins NS2B, NS4A, NS4B and characterized those interactions.

To characterize the interaction between NS3 and NS4B proteins, I performed a bacterial two-hybrid assay using either full-length or different variants (N-terminal truncations) of NS4b protein and NS3 helicase domain. The N-terminal 57 residues region of NS4B is enough to interact with NS3 helicase domain. Molecular docking of an AlphaFold2-predicted structure model of NS4B onto the crystal structure of NS3 corroborated that the N-terminal disordered region of NS4B 'wraps' around Rec-A-2 and carboxy-terminal domain (CTD) of the NS3 helicase. Molecular dynamics simulations and intrinsic tryptophan fluorescence spectroscopy studies on the NS3-NS4B complex indicated that NS4b interaction with NS3 results in dynamic motion of the CTD. Using purified recombinant NS3 helicase protein and NS4B or N-terminal 57 residues region of NS4B proteins, we showed that the helicase activity of NS3 on an RNA stem-loop substrate (a secondary structure in the viral RNA genome that must be unwound before viral RNA replication) is enhanced in the presence of NS4B.

I also characterized the essential interaction between NS3 protease domain and its cofactor NS2B protein. Crystal structures are available for a complex between a part of NS2B, that is essential for the protease activity, and NS3 proteins. I developed a novel SDS-PAGE based assay for DENV NS3 protease. Molecular docking analysis of the structure, followed by thorough biochemical analysis showed that a stretch of negative-residues in the NS2B co-factor is critical for substrate binding. Upon substrate binding, a

conformation change in NS2B, may bring the substrate bound to it, for positioning into the catalytic site for the cognate peptide bond scission as indicated by molecular dynamics simulation study and a single-molecule FRET assay results.

We also explored the interaction between NS2B3 protease and NS4A, another non-structural protein in DENV. we discovered that NS4A likely inhibits the protease activity of NS2B3, which might lead NS3 to switch its role from polyprotein processing to replication. This shift is essential for viral survival and propagation within the host cell. The findings open up new avenues for further research on how these interactions regulate DENV replication and provide valuable information for designing antiviral strategies.

Overall, this study significantly contributes to our understanding of DENV biology and the role of NS3 and its interactions with other non-structural proteins in Dengue life cycle. By elucidating these molecular mechanisms, the study provides knowledge of potential targets for drug development and paves the way for future studies that may ultimately lead to effective treatments or vaccines against DENV and other flaviviruses.

List of Figures

Introduction

Viruses as human pathogens made a notable impact on the modern human history. Yellow fever virus, a haemorrhagic (blood vessel breakage and blood leakage causing) mosquito transmitted flavivirus, is the first human viral pathogen discovered in 1900[1]. Yellow fever and malaria – two mosquito-transmitted deadly diseases – were the two main obstacles to overcome in making of the Panama canal[2]. Since then, many viruses were discovered that infect and cause a disease in humans – nearly 230 viral human pathogens are discovered so far[3]. As of July 2022, there are 129 viral pathogens as listed in the Expasy viralzone database (https://viralzone.expasy.org/678 accessed on 7 November, 2022) that are a serious concern to human health. Emergence of a novel severe acute respiratory syndrome Coronavirus (SARS-CoV-2) in late 2019 to cause the COVID-19 pandemic over the last two years is a stark reminder of the threat that viruses pose to human health.

Viruses are simple in their structural organization. In the most simplistic form, a virus is merely a supra molecular assembly of a nucleic acid, and a protein that 'coats' the nucleic acid genome. As a virus lacks cellular architecture, it exclusively depends on the cell that it infects for protein synthesis, precursors for biomolecule synthesis and energy molecules (eg. ATP) as an obligate intracellular parasite. Despite this, viruses are ubiquitously present and infect living forms from all domains of life – from bacteria to plants to insects to mammals. Unlike cellular forms of life, viruses can use either a RNA or DNA or an RNA reverse transcribed into a DNA for storing their genetic material. Based on the nucleic acid used as genetic material, viruses are categorized as RNA or DNA viruses. As cellular forms do not have a RNA-dependent RNA polymerase (RdRP), genome replication and transcription in RNA-viruses is a specialized process involving viral encoded RdRP and at unique sites in the infected cell.

From 1950 to 2017 there are nine new human viral pathogens discovered on an average every year. Interestingly, most of those newly discovered viruses are RNA viruses[3]. RNA viruses are considered a global threat, which occupy up to 44% of all emerging infectious diseases in humans[4]. Numerous RNA viruses caused epidemics and pandemic diseases since 1900. Notable of those are: Spanish flu (1918), swine flu (2009), several episodes of Ebola pandemic in Africa (2004-2014), chikungunya fever (2006-2010) and the most recent COVID-19 pandemic caused by SARS-CoV-2. Apart from these, there are several other RNA viruses that 'jumped' from other animal species into humans (zoonosis) and spreading globally rapidly. Dengue virus is one such rapidly spreading virus that affects, an estimated 390 million people annually (https://www.who.int/news-room/factsheets/detail/dengue-and-severe-dengue). The viral infection causes Dengue disease in humans and the virus is transmitted by aedes mosquitoes. The disease incidence is reported across the world. However, 70% of total disease burden is in South-East Asian countries and Africa where the aedes mosquito vector is prevalent. In the past three decades the virus and the disease is spread to the Americas and Europe where the disease incidence was not seen before due to the spread of aedes mosquito to sub-tropics and temperate regions.

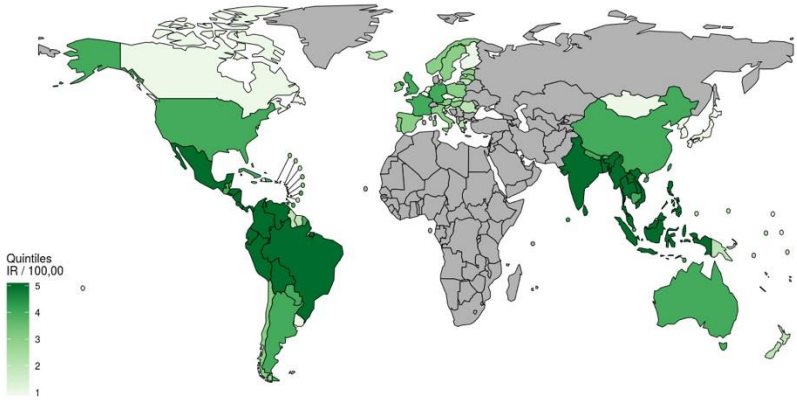


Figure 0-1 Global incidence of Dengue over the period 1990-2017. The Dengue incidence data collected from Ministries of Health of affected countries by WHO and maintained in their database is presented

chloropeth maps using the Dengue Data Application available at WHO site (https://ntdhq.shinyapps.io/dengue5/). The data from the African continent is not available in the data repository, but many African countries are badly affected by the disease

India is one of the worst affected countries by Dengue virus. Urban areas with high population density such as Delhi, Kolkata and Mumbai are badly affected by the viral infection with high incidence rates of the disease reported every year for the last 10 years. A bubble map of the Dengue incidence data over the last five years (2017-2022) is shown below. In 2021, there are 1,93,245 Dengue cases are reported in India of which 346 are fatal infections.

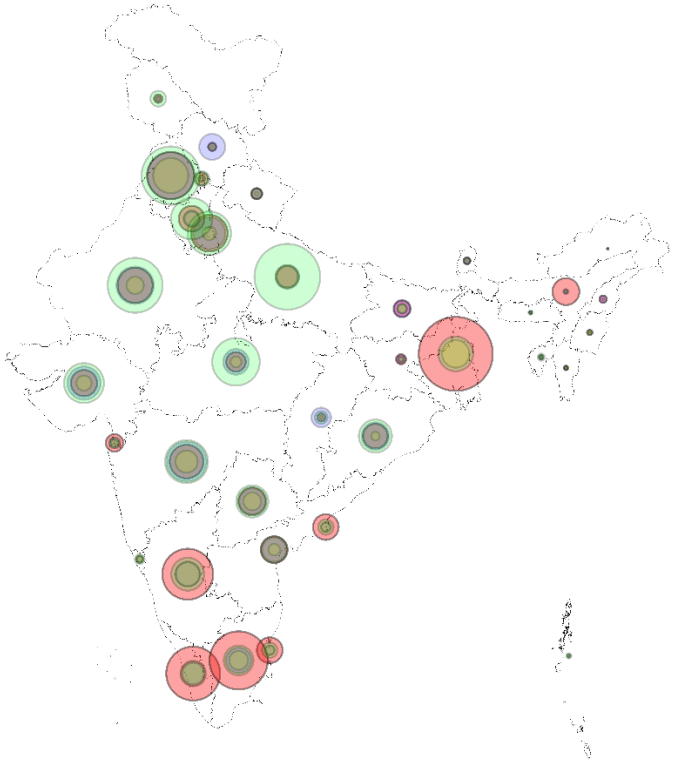


Figure 0-2 Dengue incidence in India over the period 2017-2022**.** State/union territory wise Dengue incidence data from 2017-2022 accessed from National Center for Vector Borne Disease, India website (https://nvbdcp.gov.in/index4.php?lang=1&level=0&linkid=431&lid=3715; accessed in August 2022) is plotted as bubble map on Indian geographical map using a custom python script developed in the lab using geopandas and matplotlib libraries. Different colored bubbles at the centroid of each polygon (state/UT) represent incidence in the respective state/UT for years 2017 to 2022 (till February).

Dengue Virus and Dengue Disease

The Dengue viruses are group of genetically related single-strand RNA viruses grouped under the flavivirus genus ('flavus' is a latin word for yellow color; name of the genus derived from the jaundiced - yellow color of eyes and skin - symptoms of yellow fever) in the Flaviviridae family in viral taxonomy. Other notable human pathogens in the Flaviviridae family include, yellow fever virus, tick-borne encephalitis virus, Japanese encephalitis virus, West Nile virus, the Zika virus and Hepatitis C virus (under the genus Hepacivirus). The diseases caused by these viruses are still prevalent worldwide and occasional re-emergence of these viruses in epidemic proportions (eg. Zika virus epidemic in 2016) pose a threat to global health[5].

The Dengue viruses (DENVs) are a group of genetically related but antigenically distinct viruses that are known as serotype 1-4 (DENV-1, -2, -3 and DENV-4). DENV is thought to have evolved as a non-human primate virus transmitted by tree hole mosquitoes amongst the animals before a zoonosis event into human host. After the zoonosis (probably in Asia) the virus would have diverged into four distinct serotypes (DENV 14) and circulated among the human population through Aedes aegypti mosquito vector transmission[6]. Each serotype has significant genomic divergence. Serotype-1 (DENV-1) and serotype 2 (DENV-2) are more prevalent in India as per the seroprevalence data[7]. Infection by any of the serotype of the virus causes Dengue disease.

Similar to other Flaviviridae family members, Dengue virus has a single stranded RNA genome that is 10.72 Kb (NCBI Reference Sequence: NC_001474.2) nucleotide long. The genomic RNA is a positive-sense strand (with respect to translation) that gets directly translated into protein in the infected cell. There is a single ORF in the sequence, flanked by a 5'-UTR and 3'-UTR. The viral RNA is 5'-capped. More detailed description of the

genome structure and organization is presented in later sections of this chapter. The virion is spherical in shape with 250Å radius. The capsid (the protein shell covering and protecting the genomic RNA) is a icosahedron (T=3) shell made from the capsid (C) protein. The virus has a lipid bilayer envelope covering the icosahedral capsid and the genome inside it. The viral envelope protein, E, is tail-anchored into the membrane and arranged on the virion surface in a well-organized pattern - the characteristic 'herringbone' arrangement of E-dimers on the virion surface. Apart from capsid and E protein, the viral structure also has M protein ('M' for membrane protein; 38 amino acids long; formed from a precursor protein, prM, after processing by a Furin in the Golgi vesicles) embedded in the viral membrane. Altogether, capsid, E and M proteins (M is proteolytic product of a precursor protein, prM) along with the genomic RNA and the membrane make up the whole virion structure[8].

Infection with DENV causes a febrile disease caused dengue. The disease manifestation is very complex. The disease severity varies from mild to very severe. The mild manifestation of the disease is dengue fever. A sudden onset of high grade fever (102103 ^{o}F) accompanied with nausea/vomiting, rash/red patches appearance on skin of limbs, severe muscle and join aches and pain, pain in eye orbit are characteristic symptoms of dengue fever. Quite often these symptoms are mistaken for other viral infections such as flu. In few cases (nearly 5% of cases), the disease can progress to a severe form the disease called dengue haemorrhagic fever (DHF). As the name suggests, DHF is characterized by symptoms of severe internal bleeding due to rupture and leakage of blood vessels and capillaries with clinical accumulation of fluid, persistent vomiting and mucosal bleeding. DHF severity is reported in four grades, and grade III and IV are marked as the onset of dengue shock syndrome (DSS) – the most difficult phase of the disease to clinically

manage. DSS is characterized by multiple organ failure, respiratory distress, severe liver enlargement due to hypovolemic shock caused by severe lymph leakage into visceral cavity and into lungs. As per WHO guidelines on Dengue, published in 2009[9] the disease is graded either as Dengue with or without warning signs, or as severe dengue (where hypovolemic shock [DSS] is included).

Dengue Vaccine, Drugs, and Current Therapeutic Strategies

There are no affective vaccines against Dengue. Currently, there is only a single licensed vaccine (Dengvaxia® ; developed by Sanofi Pasteur) recommended for use in high disease prevalence regions. However, the vaccine efficacy against the serotypes 1 and 2 is only ~50% as per the phase III clinical trials[10]. There are no small molecules drugs against Dengue. The current treatment regimen is mostly for symptomatic relief. There is an urgent need for affective drug against the viral infection. Targeting the viral proteins involved in viral genome replication and protein processing potentially can be used as an antiviral strategy. Several small molecules that block the viral protease and/or interactions amongst the viral proteins required for viral genome replication have been tested. Recent trails of a small molecule drug, JNJ-A07, that blocks interaction between NS4b and NS3 proteins, essential in viral RNA replication, showed significant results in clinical trials[11]. However, it is important to note that a detailed characterization of the structure and conformational dynamics of these two proteins and their interaction is lacking. The work embodied in this thesis was planned to explore the interactions between the DENV replication proteins and understand the functional significance of the interaction.

Dengue Life Cycle

Dengue virus (DENV) is transmitted to humans by the mosquitoes Aedes aegypti and Aedes albopictus. When ingested into the mosquito midgut with blood, DENV first

interacts with midgut cell membrane receptors and then enters the cells through receptor-mediated endocytosis. Following replication in the midgut, DENV disseminates to salivary glands for transmission to human.

Cellular Tropism and Spread Inside the Cell

DENV entry into the body through the skin by a mosquito bite. Cells with the replicative form of DENV were observed in the skin (dermal dendritic cells, Langerhans cells, Monocytes derived dendritic cells, and dermal macrophages), liver (hepatocytes), and spleen (Macrophages in the red pulp, splenic sinusoidal endothelium)[12–17].

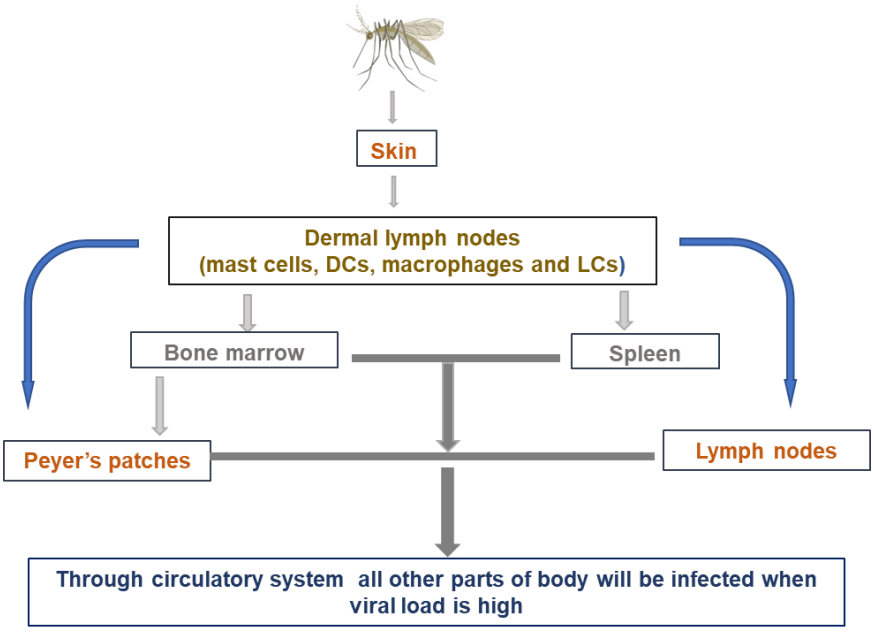


Figure 0-3 DENV entry into the body and its spread to different organs

Viral Entry into the Cell and its Receptors

DENV chooses multiple pathways for entry into the cell, a multidomain protein on the envelope of DENV helps for this purpose called E protein (Envelope protein), along with pr and M proteins. Domain III of E protein involved in attachment, cellular recognition

internalization of DENV[18]. DENV uses several attachment factors and receptors to enter inside a cell, majorly Heparin sulphate, proteoglycans, dendritic cell-specific ICAM3grabbing nonintegrin (DC-SIGN), lymph node-specific ICAM3 grabbing nonintegrin (LSIGN), mannose receptor, Hsp90, CD14, GRP78/BiP, and a 37/67-kDa high-affinity laminin receptors [19]. The outer leaflet of a viral envelope containing phosphatidyl-serine and phosphatidyl-ethanolamine lipids, recognized by TIM/TAM family receptors, helps

in viral entry to the cells by viral apoptotic mimicry mechanism[20,21]. After receptor recognition, DENV slides into already preformed clathrin-coated pits; internalization happens mainly through clathrin-mediated endocytosis; sometimes, DENV chooses noncanonical endocytic pathways independent of caveolin, clathrin, and lipid rafts. The process of DENV internalization depends on the type of cell which DENV is infecting and the serotype of DENV[22,23]. Later in late endosome membrane fusion, the release of genomic RNA into the cytosol happens (Figure 0-4) .

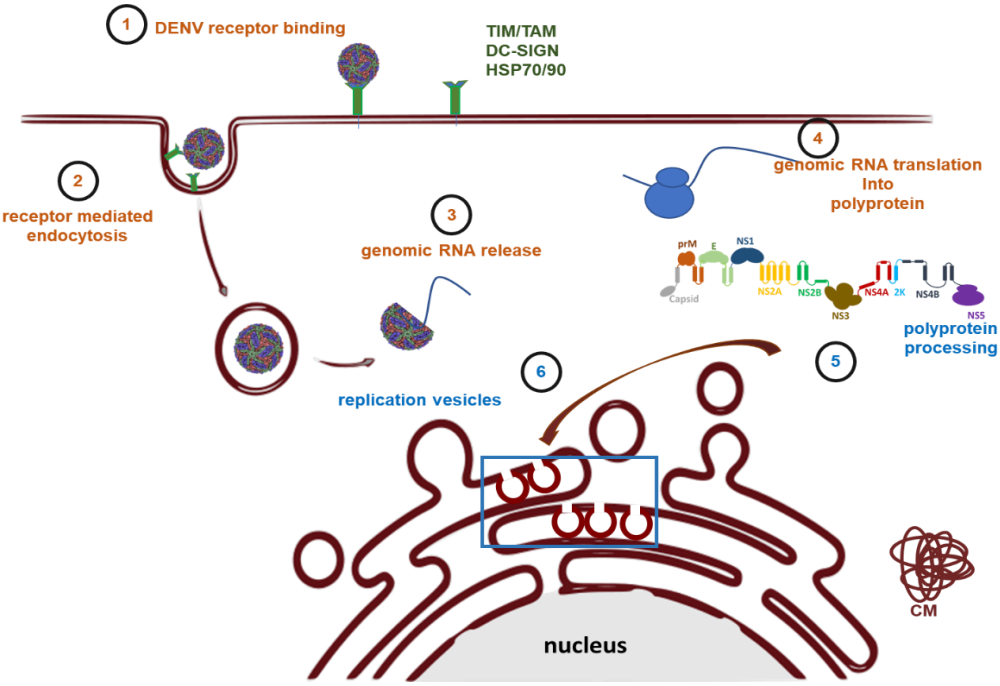


Figure 0-4 DENV entry and different stages of DENV replication inside the infected cell

Dengue Virus Genome Organization

Upon entry into the cell, genomic RNA of DENV is released into the cytosol for translation then transcription/replication happens inside the cytosol. The viral RNA is translated first, and then viral proteins synthesize a complimentary strand of the genomic RNA (c-RNA or negative-strand) in specialized membrane-bound complexes. The negative-strand RNA in-turn serves as a template for synthesis of new viral genomic RNA copies. The viral RNA has a single ORF and by translation produces a single long polypeptide, called polyprotein. The polyprotein is processed into individual proteins by host and viral encoded proteases to give individual functional proteins[24]. The genomic RNA of DENV has a 5'-cap, but there is no 3'- poly-A tail. The single ORF in the genome is flanked by 5'- and 3'- untranslated regions (UTRs). Both 5'- and 3'- UTR sequences have extensive secondary structures and have relatively conserved sequences among all flaviviruses (Figure 0-5A). The genome undergoes a cyclization event inside the host cell to generate a "pan handle" (Figure 0-5B) structure. Cyclization of the RNA happens through long-range RNA-RNA interaction by 5'-CS (cyclization sequence) and 3'-CS, which are present in 5'- and 3'- UTR regions. In the 5'- UTR, some parts of the capsid sequence also participate in cyclization. The 5'- and 3'- CS have ten or more contiguous nucleotides that perfectly complement all mosquito-borne flaviviruses CS sequences[25]. Circularized panhandle structure having RNA genome maintain some of the secondary structures after cyclization, which plays an essential role in RNA replication and translation[26,27].

5'-UTR The 5'-UTR contains two major secondary structural domains; the first domain contains a Y-shaped stem-loop structure called SLA. This structure is conserved among

flaviviruses, despite that there is no sequence conservation[28]. The intact SLA structure is needed for the proper replication of the DENV genome, and it interacts with viral RNA dependent RNA polymerase (RDRP or NS5)[29]. Another highly conserved secondary structure (among several flaviviruses) in the 5'-UTR is SLB, which contains 5'-CS[30]. SLA and SLB regions in the 5'-UTR acts as promoter elements for

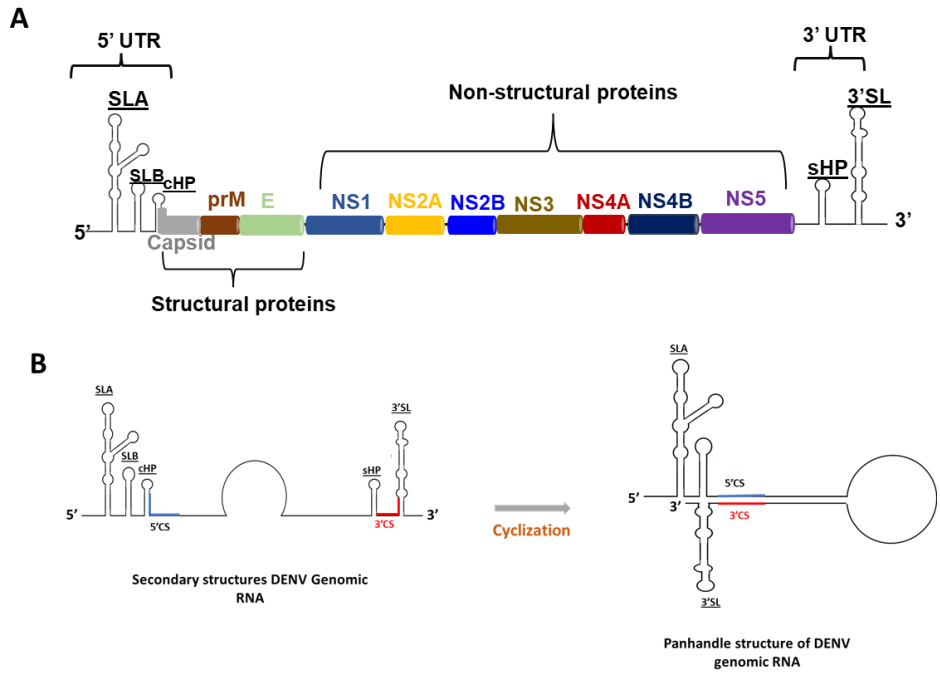


Figure 0-5 DENV genome-(A) Genome secondary structures with single ORF coding polyprotein, (B) DENV genome cyclization

3' UTR The 3′-end of the positive-strand genomic RNA is the starting point for negative strand RNA Synthesis[31]. 3'-UTR has several secondary structures. These structures do not significantly affect the translation of the viral RNA, but seriously compromised or abolished new RNA synthesis when disturbed. The complementarity between sequences at the 5'-CS and 3'-CS ends of the genome is essential for dengue virus RNA synthesis. Deletions in this region resulted in replicons with decreased RNA amplification[32]. The 3'SL region at the end of 3'-UTR is essential for negative-sense RNA synthesis; the

unwinding of this RNA stem-loop structure leads to the synthesis of negative-strand RNA[33]. 3'-UTR interacts with viral coded NS2A and 4A and is essential for viral-induced membranous structures on ER in a non-replicative manner[34,35]. Along with these secondary structures in 3'- and 5'- UTR of the genomic RNA coding region of the genome also has specific secondary structures; RNA hairpin structure in the capsid coding region (cHP) directs translation start site selection in human and mosquito [36]. More detailed description of the secondary structural elements in the 3'-UTR, its interaction with viral replication proteins and its significance in viral RNA replication is presented in the introduction of chapter 2.

Translation of DENV Genomic RNA

After virion entry into the target cell, the genome is released into the cytosol by capsid uncoating; this process may be done through ubiquitination of capsid protein [37]. First round translation of DENV genomic RNA occurs in the cytosol at the surface of the ER to produce a single polyprotein which is further processed co- and post-translationally by host and cellular proteases to give all functional viral proteins. Genomic RNA of DENV contains a type-1 5'-cap structure (m7GpppAm2), which enables translation through canonical cap-dependent translation initiation[38]. Poly-A-binding protein which interacts with one of the translation initiation factors, eIF4E, mediates the circularization and initiation of translation in eukaryote cells, interacts with the 3'-end of DENV RNA, and initiates translation (3'-SL region alone influence the translation after 40S ribosomal subunit binding) [39,40]. In flavivirus 5'- region has two start codons; translation initiation from the correct start codon is influenced by a stem structure in capsid coding region called capsid coding region hairpin (cHP). Mutations in this region disturb the secondary structure and decrease initiation from the first AUG[41]. Cap-independent translation of

RNA is also described for DENV; although there are no IRES elements like Hepatits C virus, this cap-independent translation is majorly regulated by 3'- and 5'- UTRs [42]. In the host cell, DENV induces translation inhibition at the initiation step and is not functionally linked to PKR, eIF-2 phosphorylation of stress granule (SG) induction. Through specific unknown mechanisms, DENV manipulates host cell gene expression at the translational level.

The single ORF flanked by 5'- and 3'- UTRs codes for the 11 proteins as a single polyprotein this polyprtoein is processed into individual functional proteins by host and viral coded proteases. These processed proteins are classified into two categories structural and nonstructural proteins:

Structural Proteins Capsid, pr, M, and Envelope protein - these proteins make the structure of the virus particle. Protein M is seen only in the mature virion after budding out of the host cell, after proteolytic processing of its precursor protein, prM.

Nonstructural Proteins (NSs) there are seven non-structural proteins that result from proteolytic processing of the polyprotein: NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5, and a small peptide, 2k peptide, present in between NS4A and NS4B. These proteins are called non-structural proteins as these proteins are not part of the final virion structure. Non-structural proteins help the virus for its polyprotein processing, replication, genome packaging, and virus particle synthesis. Out of these 7 proteins NS3 (protease, NTPase and helicase) and NS5 (methyl transferase and RNA dependent RNA polymerase) are enzymes remaining proteins are predicted transmembrane proteins with some portion of their structure being cytosolic. Structure and functional characterization of the

nonstructural proteins and understanding the complex and dynamic interactions among those during viral replication is the focus of this thesis study.

Dengue Virus (RNA) Replication

Dengue viral replication or multiplication inside the infected cells happens in 5 steps:

Step 1: polyprotein formation through the translation of genomic RNA

Step 2: polyproteins processing and replication vesicle formation,

Step 3: negative sense RNA synthesis by the RNA dependent RNA polymerase (RdRP, NS5) from the 3'-UTR promoter.

Step 4: 'strand switching' by the RdRP for replication of the positive sense RNA using the negative-strand RNA as a template.

Step 5, virus particle synthesis by packing of positive sense RNA with the capsid

From the polyprotein processing step to the virus particle assembly step, non-structural proteins will make complex and dynamic interactions with each other and host proteins. Interaction between the non-structural proteins is critical for viral multiplication in the cell. During the negative-strand synthesis and viral RNA synthesis, double-stranded (ds) RNA intermediates are formed, which can potentially trigger dsRNA-sensing-and-decay mechanisms of the host cell. Non-structural proteins' interaction with host cell proteins and modulation of the membrane structures of the host cell are a key aspect in viral replication.

Polyprotein Processing

The viral polyprotein synthesis happens on rough ER surfaces. The newly made polyprotein, as it exists from the ribosome, is translocated into the ER membrane. As rest if the polyprotein is translated, a major part of the polypeptide inserts into the ER membrane and fold into transmembrane domains. The initial polyprotein cleavage events

happen at the ER surface. After that, the partially processed NS proteins are transported to virus-induced convoluted membranes[43]. Though it is not completely understood, several host proteins and viral NS proteins (such as NS4a) are implicated in inducing curvatures in the ER membrane such that ER membrane folds into the lumen of the ER. The polypeptide is cleaved at specific sequences recognized by viral coded proteases and sometimes host proteases to give rise to individual functional proteins essential for virus survival and amplification inside the infected cell.

One of the NS proteins, NS3, has serine protease activity. The N-terminal one-third of the NS3 protein has a canonical chymotrypsin fold, and along with its essential co-factor protein, NS2b, makes proteolytic cuts at specific NS protein boundaries to form different NS proteins (Figure 0-6).

The NS2B cofactor - NS3 protein complex is a two-component protease complex that cleaves at the specific site containing one or more 'basic residues' ('K/R') and a small side chain having amino acid (mostly S/G/V) in the scissile site. Along with NS2B3 protease, host signalase, and Golgi protease also perform polyprotein processing[44]. After polyprotein processing, all the functional nonstructural proteins will be involved in the formation of a replication vesicle on the ER membrane where genomic RNA replication happens[45]. In other RNA viruses like Nora Virus, this switching from polyprotein processing to replication happens as a spatial and temporal event with dynamic interactions of nonstructural proteins mainly through viral coded protease [46].

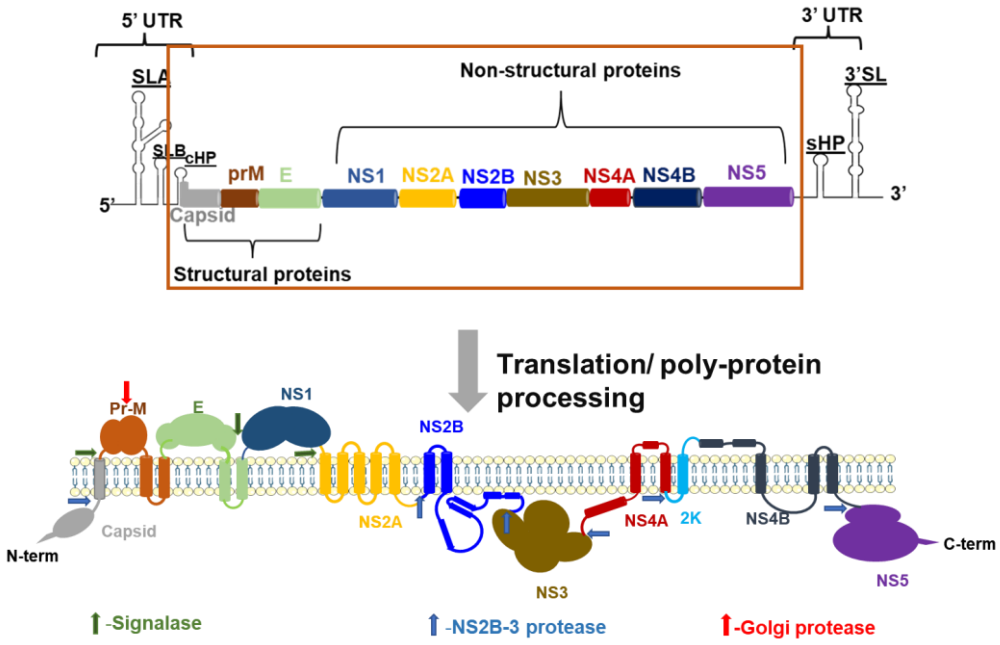


Figure 0-6 ORF of genomic RNA coding for the polyprotein and poly protein processing by viral and host-coded protease

The Viral Replication Organelles/Replication Compartments

Positive-sense RNA viruses induce extensive membrane rearrangements in the cell (Flaviviridae-ER, Bromoviridae-ER, Nodaviridae-Mitochondria, Togaviridae Plasma membrane/endosomes/lysosomes, Tombusviridae-peroxisomes). These virusspecific membrane rearrangements include single or double-membrane vesicle invaginations and convoluted membranes[47,48]. Flaviviruses are well-known for extensively remodeling different host organelle membranes to facilitate efficient viral RNA replication. These membranous structures give certain advantages to the virus by spatially partitioning RNA replication from viral particle assembly and enriching all the proteins at a confined location with other metabolites, and protecting all viral components from host innate immune responses and degradation, giving the virus an efficient way to do replicate its genome and virus particle synthesis[49]. Electron tomography studies of

DENV infected hepatoma cells showed single membrane invaginations into ER lumen with a vesicles type of structure with ~90nm size and some unstructured convoluted membranes (CMs); CMs of DENV infected cells are rich with all nonstructural proteins. However, no double-stranded RNA in convoluted membranes indicates this may be a site of RNA translation and polyprotein processing[45], and a recent study showed CMs involved in regulating nonstructural protein levels in the cells through ER-mediated degradation mechanism [43]. These vesicle-like invaginations are observed in ER of mosquito cells infected with DENV, but they lack CMs [50]. NS4B induces DRP1-mediated elongation of mitochondria and affects the biogenesis of CMs in dengue-infected cells, which enhances viral replication and impaired immune responses inside the cell[51]. On the other side, vesicle pockets in DENV-infected cells colocalized with dsRNA and other nonstructural protein complexes indicated that these inner vesicles are the likely sites of genome replication.

Furthermore, these are juxta positioned near virion assembly sites and connected with a pore-like opening. And these vesicles are also called 'replication vesicles,' and nonstructural proteins inside these vesicles form complexes through interacting with each other and some host proteins and are involved in the replication of viral genome called 'replication complexes (RCs)' [45,52]. DENV-induced vesicle pockets have similar morphology to their close relatives WNV, ZIKV, and TBEV; these structures are evolutionarily conserved among flavivirus and may share a similar replication strategy for their genomes [53–55]. In DENV infected cells, the biogenesis of the replication vesicles and convoluted membranes was induced maybe by NS4A mediated autophagy or NS3 recruits fatty acid synthase to this replication complex and activates it, which may help

the synthesis of lipids which in turn helps in vesicle formation[56–58]. Along with other nonstructural proteins 3'-SL region of DENV 3'UTR also has some role in DENV vesicle pocket formation [35].

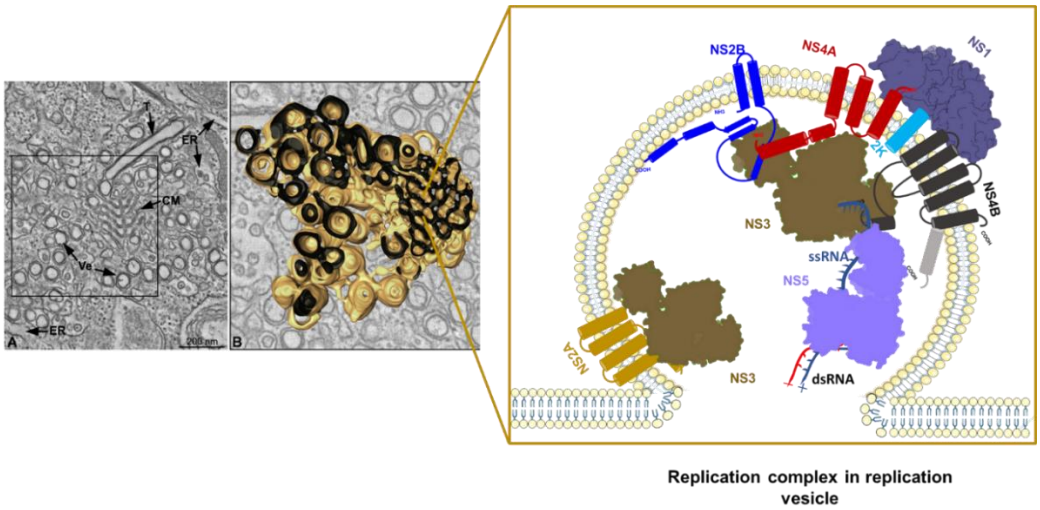


Figure 0-7 Electron microscopy tomography image (tomography image from ref. 59) of membranous replication vesicles seen in DENV-infected cells with a schematic of replication complex inside replication vesicle

The RC is a multi-component structure consisting of viral proteins, host proteins, and the RNA template and dsRNA intermediate, assembled for the replication of viral RNA, using colocalization with dsRNA as a marker for RNA replication, deletion, and mutagenesis studies of individual nonstructural proteins. Nonstructural proteins have been reported to play a role in the viral RNA replication [45,48,52,55]. However, the exact stoichiometry and architecture of the Dengue RC remain unknown.

Model for Viral Replication

Based on the events and interaction interface of the nonstructural and their little-known functional importance led to the creation of a model for replication events in the dengue induced replication vesicles. After multiple translation events inside the cytosol, nonstructural proteins make replication vesicles inside the cell on the ER membrane.

Where circularized genomic RNA comes, start synthesizing new copies of the genome. RNA replication is a highly complex and tightly regulated process that requires the recruitment of host proteins (also known as host factors) are critical for proper RNA replication[60].

The two critical proteins in viral genomic replication are two enzymes: NS3 and NS5; the viral helicase-cum-protease and RdRP, respectively. Role of other non-structural proteins (NS2A, 4A and 4B) is not understood, though those proteins are critical for replication. In a proposed mechanism, other nonstructural proteins (other than NS3 and NS5) interact and modulate the function of NS3 and NS5 protein in different genome replication steps. However, understanding on how NS4A, NS4b, NS2A interact with NS3 and/or NS5 and the functional significance of these interactions in viral RNA replication is limited.

After replication vesicle formation, all nonstructural proteins, along with viral RNA and dsRNA replication intermediates localize in the replication vesicle[59], as schematically depicted in the Figure 0-7. The initial recognition of genomic RNA happens through multiple interactions with non-structural proteins such as NS3 (C-terminal domain), NS4A, and NS2A. Furthermore, the interaction of NS3 with NS4A, NS4B, and NS5 stimulates and enhances its unwinding activity on long RNAs forming a panhandle structure and other secondary structures to linearize genomic RNA for replication. After linearization, genomic RNA will be handed over to NS5 - the main enzyme involved in synthesizing the new negative-sense RNA genome using viral positive sense RNA genome as a template - to form a dsRNA intermediate. From this negative-sense RNA genome is subsequently used as a template for the synthesis of the additional positive sense RNA genome, and an asymmetric replication occurs, and approximately 10 to 100 times more positive sense RNA than negative sense RNA is synthesized by NS5 RdRp

during replication [61,62]. This asymmetric replication may happen through two different complexes inside the replication vesicle with and without NS2A. The complex without NS2A will do the synthesis dsRNA intermediate, and the complex with NS2A will do the unwinding of this dsRNA and synthesize and send out positive sense RNA for packaging into the structural proteins with the help of NS3 by its interaction with structural protein. This leads to the accumulation of more negative sense RNA inside the replication vesicle, causing the synthesis of more positive sense RNA. Thus, understanding structure function relationship of different NS proteins and their interactions with other NS proteins and the viral RNA is important for explaining the replication mechanism of DENV and other flaviviruses. Knowledge of the interactions and the functional significance of those interactions will help greatly to develop novel therapeutic drugs to prevent viral replication – such as JNJ-A07 described earlier.

Structure-Function Relationship of Dengue Virus Non-Structural Proteins

There are seven nonstructural (NS) proteins (NS1, NS2a, NS2b, NS3, NS4a, NS4B, and NS5). These nonstructural proteins are proteolytically processed and present only in DENV-infected cells. These nonstructural proteins have distinct functions inside the host cell, including defence against the host immune system, replication, and viral genome translation.

NS1: is a monomeric protein with molecular weight ranging from 46 to 55 kDa depending on its glycosylation and exists in two forms: membrane interacting dimer and secretory lipoprotein of the hexamer. NS1 colocalizes with double-stranded RNA and membranous structures induced by the virus on ER, indicating its role in replication. Pre-exposure of NS1 to dendritic cells showed enhancement in replication. NS1 has a crucial role in early replication events, especially negative sense RNA synthesis[63–65]. Some NS1 protein

interactions with host factors are also essential for replication. In DENV, NS1 protein interaction with the ribosomal protein RPL18 is required for viral translation and replication[66].

NS1 monomer has three structural domains (Figure 0-8), namely a β-roll dimerization domain(**1-19aa**), a wing domain (**30-180aa** with two glycosylation sites **N130, N175**), and a β-ladder domain (**181-352aa**) at C-terminus [67,68]. Soluble monomer dimerizes to gain partial hydrophobicity, allowing it for membrane association, and this functional dimeric form of NS1 is essential for viral replication. During polyprotein synthesis, NS1 is translocated into ER lumen co-translationally by the transmembrane signal peptide present at the C-terminus of E protein and coats ER-derived membrane vesicles from the lumen side as a dimer[69,70]. Comparative analysis of DENV and WNV with ZIKV NS1 showed a crucial role in membrane remodelling, a region ranging from 114-129 amino acids of wing domain loop spike has characteristics of fusion loop and interact with ER membrane; this interaction requires a functional dimer. It is likely that through this membrane remodelling function NS1 effects viral replication by without interacting with RNA or other enzymes involved in replication[68,71,72].

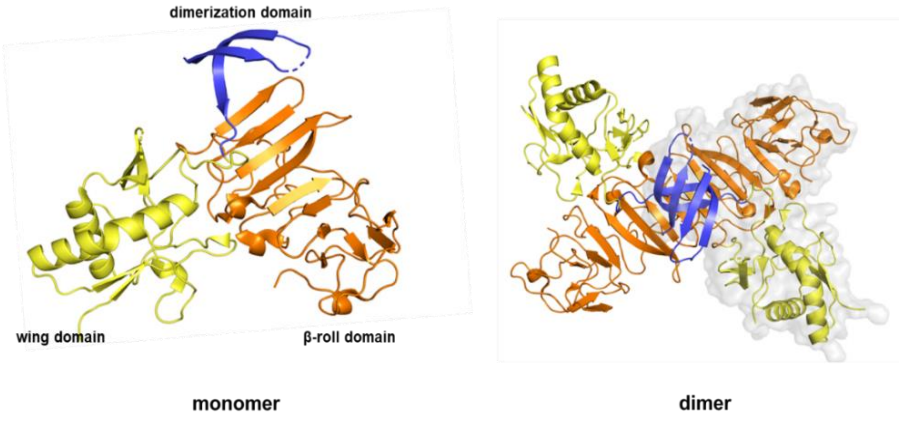


Figure 0-8 Nonstructural protein-1 in monomer and dimer form with domain labelling (PDB ID

Secretory NS1 (sNS1) is a barrel-shaped hexameric lipoprotein with a central opening, and secretion of sNS1 depends on its glycosylation status. NS1 undergoes posttranslational modifications such as glycosylation at N135, N207, and N175 in the trans Golgi network of the host; in some flaviviruses, at least one of the two N-glycosylation sites (probably N135) in NS1 protein is essential to produce viable virus.

NS2A: NS2A is a 22kDa predicted transmembrane protein, processed by both host signalase at N-term between NS1 and NS2A and C-terminus between NS2A and NS2B by viral coded protease[73,74]. Topological studies of NS2A predicted there would be eight predicted transmembrane helices known to associate with the endoplasmic reticulum (ER) membrane. Out of these, N-terminal 68 amino acids are located in the ER lumen, amino acids 69 to 209 form five transmembrane segments (Figure 0-9), and the C-terminal tail (amino acids 210 to 218) is located in the cytosol region [75]. The role of NS2A in virus particle synthesis is first reported in the Kunjin virus; mutations at positions 59 and 149 in the Kunjin virus NS2A affected virus particle production[76]. In another study, a point mutation in dengue NS2A R84E showed an abortive replication which indicates NS2A is also essential for viral replication[75]. Two different forms of NS2A are observed, in which one helps in replication, and the other helps in viral nucleocapsid formation[77].

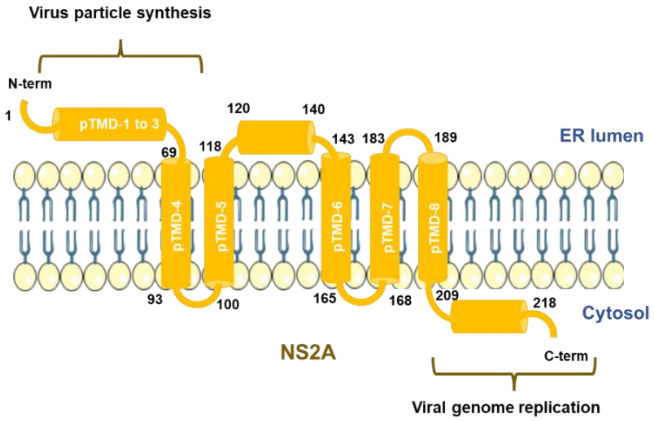


Figure 0-9 NS2A protein membrane topology model with predicted trans membrane helices

Along with replication and virus particle synthesis, in other flaviviruses, NS2A has a role in inhibiting IFN β- promoter-driven transcription; specific point mutations like A30P showed reduced inhibitory effect and lesser neurovirulence in WNV and also induced apoptosis which is independent of interferon response[78–80].

NS2B: NS2B is a 14kDa protein processed on both sides by viral coded protease[81]. With predicted three possible transmembrane helices and an internal ~50 amino acids (45103aminoacids) stretch of cytosolic region (Figure 0-10) which interact with NS3 and act as a cofactor for its protease domain. NS2B may be incorporated into ER membrane co-translationally, and this membrane interaction enhances cleavage activity between NS2B and NS3[82,83]. Along with its cofactor role in the NS3 protease domain, it was shown to form a homomeric oligomer; the remaining predicted transmembrane regions have a possible viroporin function and structure and have a role in recruiting NS3 at ER-derived replication vesicles[84–89]. Systematic mutagenesis at the flavivirus conserved residues within the transmembrane domains of JEV NS2B showed it is essential for viral RNA synthesis and Virus particle formation[90].

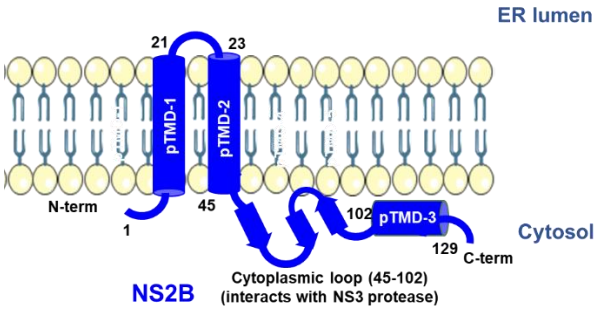


Figure 0-10 NS2B protein membrane topology model with predicted transmembrane helices and soluble cytoplasmic loop

NS3: NS3 is a 69kDa soluble cytosolic multidomain multifunctional protein with two different functions in two different subdomains [81,91–93]. The N-terminal part of NS3 is a serine protease, and the C-terminal part is an SF2 family helicase with a C-terminal accessory domain.

Protease Domain (NS3pro) consists of N-terminal 175 amino acids and is a trypsin-like serine protease cleaves after a 'basic' amino acid: arg/lys, mostly arginine, and most of the polyprotein processing of nonstructural proteins and between prM/Capsid. NS3pro recognizes a specific sequence pattern between junctions of nonstructural proteins[94]. NS3pro adopts a chymotrypsin-like fold (Figure 0-11), with the catalytic triad "His50Asp75-Ser136" [95,96]. Structurally protease domain exists as an independent domain away from the core helicase domain with a flexible loop. As mentioned previously, the cytoplasmic region of NS2B acts as a cofactor for the NS3 protein, and this interaction of NS2B brings it to ER membrane, where it will do proteolytic processing of nonstructural protein. Along with protease activity, this interaction of NS2B to the NS3 protease domain increases NS3 helicase domain nucleic-acid binding specificity towards RNA[97].

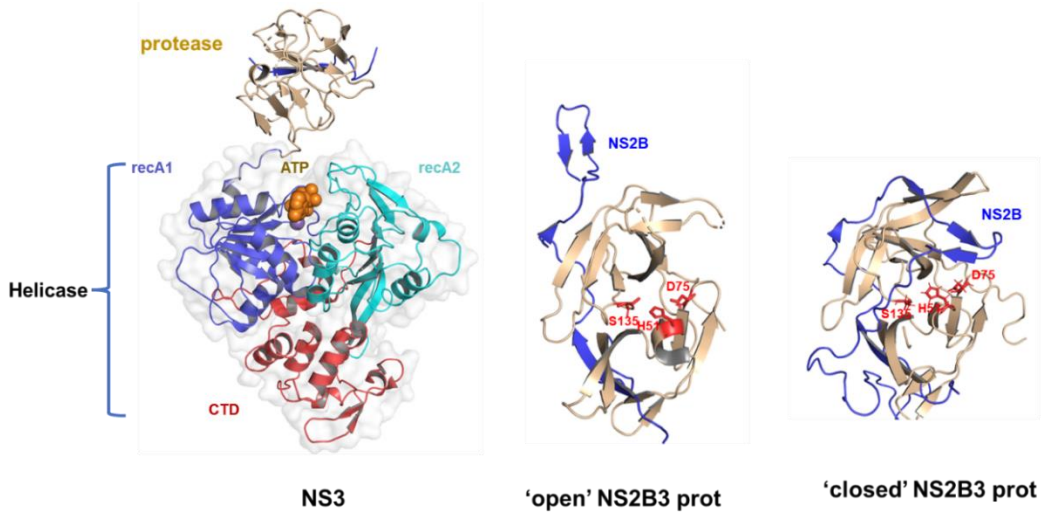


Figure 0-11 NS3 protein domain architecture and NS3 protease two different conformations with its NS2B cofactor region (PDB ID- 3L6p (open), 2m9q (closed))

The Helicase Domain (NS3hel) consists of remaining residues with two recA domains and an extra C-terminal domain with a different fold in flavivirus compared to HCV of the same Flaviviridae family[98,99]. NS3 helicase belongs to the NS3/NPH-II family of SF2 superfamily with eight conserved motifs (Figure 0-12)- Motif I (193-205), Motif Ia (222235), Motif II (280-295), Motif III (310-320), Motif IV (360-375), Motif IVa (385-395), Motif V (407-420), Motif VI (450-460). These motifs have a specific function in helicase activity; motifs I, II, and VI are involved in NTP binding and hydrolysis; Motif Ia, IV, IVa, and V are involved in nucleic acid-binding. Motif III does the coupling of ATPase and unwinding of bound nucleic acid polymer, and Motif V does energy transduction between ATPase and RNA binding cleft. This family of helicases has RNA or DNA stimulated nucleoside 5'-triphosphatase (NTPase) activity (mostly ATPases), RNA 5'triphosphatase (ATPase) activity, and helicases activity with 3'to 5' directionality; all these activities are divalent cation (mostly Mg^{+2}) dependent, and both rec domains and c-terminal domain are essential for this activities[95,100–105].

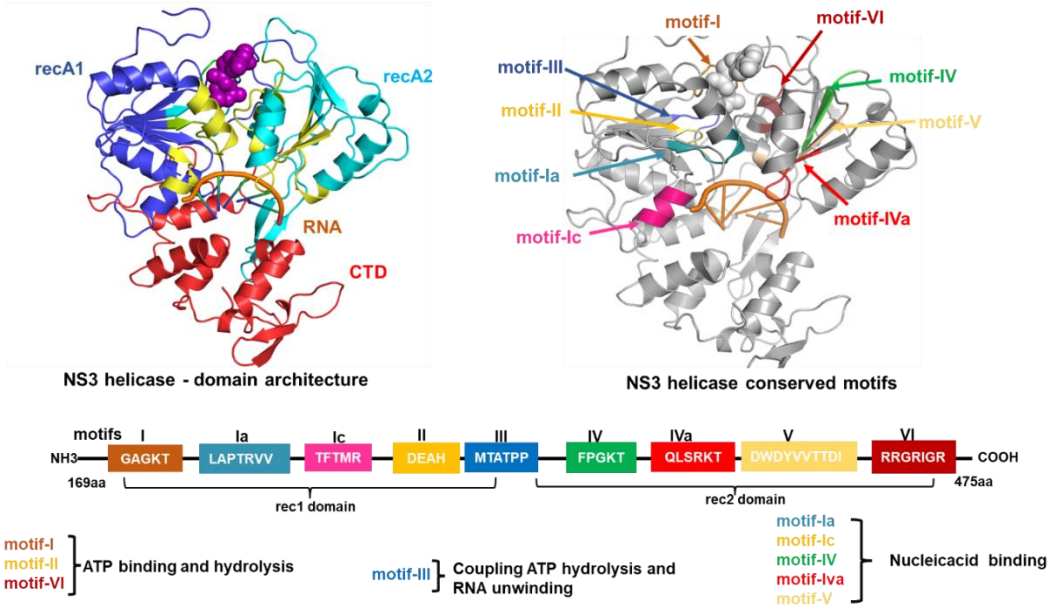


Figure 0-12 ssRNA and ADP bound form of NS3 helicase domain of NS3 with subdomain and Motif marking (PDB ID- 2JLX)

DENV helicase recognizes a conserved 12-mer of 5'-genomic RNA region, which stimulates RNA-mediated ATPase activity and indicates NS3's specificity towards viral RNA. Amino acids from both the Rec-A domains of the helicase domain are involved in this specific interaction (D290 and R538)[106]. Asp 290, which has initial interaction with RNA, is next to the DEXH motif and interacts with R387 in Motif IVa, which involves inserting RNA into the helicase core, this induces inter-domain closure which changes the divalent metal ion coordination, which activates ATPase activity and RNA binding mediated ATP hydrolysis [107]. Along with this helicase activity, NS3 has ATP independent RNA annealing activity, promotes RNA secondary structure, and dsRNA formation [108]. NS3 also interacts with FAS (Fatty acid synthase) and recruit this protein at the replication vesicle site. This interaction is essential for the replication of vesicle formation and expansion[58]. A proline-rich region in between the protease and the helicase domain (towards N-terminus of the helicase) is essential for virus particle synthesis.

Mutation of this region showed lesser infectivity than wild type[109].

NS4A: NS4A is a 14kDa trans-membrane protein processed by both viral and host proteases as an initial 4A-2K-4B intermediate. Later it will be further processed into a mature NS4A protein, and a small portion of NS4A/4B cleavage happens post-translationally, N- terminal cleavage between NS3/NS4A done by NS3 protease C-terminal cleavage between NS4A and 2K mediated by host protease which has similar properties like signalase [110]. It is an alpha-helical protein with three transmembrane helices (Figure 0-13) ranging between 56-75, 81-100, and 104-120 amino acids. NS4A is essential for viral replication when expressed in mammalian cells without 2K peptide-induced cytoplasmic membrane alterations, which are comparable to virus-induced ER membrane alterations; when there is a 2K peptide, there are no significant membrane alterations in ER membrane which indicates NS4A helps to induce ER membrane vesicle formation in a 2K regulated manner[111–113].

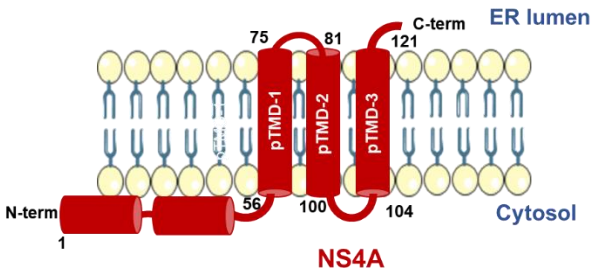


Figure 0-13 NS4A predicted transmembrane topology with transmembrane helices

N-terminal amphipathic helix of NS4A involved in homomeric oligomerization and exact oligomeric status is not yet known, and this amphipathic helix is shown to be essential for viral replication in DENV. When considered and compared with another flavivirus like WNV, the N-terminal amphipathic helix region is essential for NS4A stability and proper folding, and this same region showed to interact with curved lipid membranes [114– 117].

Along with this membrane modulation and role in the replication of DENV, it has several other functions like in JEV NS4A showed interaction with DDX42 RNA helicase and subsequently modulated interferon response to void the immune response of host for viral survival inside the cell[118]. Moreover, in DENV and ZIKV, it induces autophagy in the infected cell; this induction of autophagy also fuels replication vesicle formation by inducing lipid droplets lysis, which helps the virus to sustain inside the cells and enhance its numbers[56,119].

NS4B: NS4B is 27kDa trans-membrane protein first produced as a 30kDa precursor protein with a 2k signal peptide and then post-translationally modified into a 27kDa mature NS4B C-terminal cleavage between NS4B/NS5 mediated by NS3 protease[120]. NS4B is initially processed as NS4A-2K-NS4B precursor, which is essential for viral replication and ER localization of this protein which is mediated by 2K signal [121]. NS4B protein is highly conserved among the DENV serotypes, with an overall protein sequence conservation of 78%. NS4B interacts with most of the host proteins; these interactions are essential for viral replication and survival inside the cell by avoiding innate immune response (NS4B inhibits STAT-1, inhibits Dicer activity, NS4B activates unfolded protein response by

ATF6/IRE-1 pathway, and helps to overcome ER regulated stress by interacting with SERP1)[122,123]. Among all nonstructural proteins, this is the only protein with unknown function with inhibitors such as NITD-618, SDM25N, and AZD0530 that will bind to it and stop viral progress inside the cell [124]. NS4B, apart from ER, is shown to colocalize in the nucleus, mitochondria, and phagosomes. Although the exact function of colocalization in the nucleus is unknown, NS4B induced morpho-dynamic changes in

mitochondria (elongation of mitochondria) through inhibition of DRP1, which promoted DENV replication by protecting host immune response [51,125].

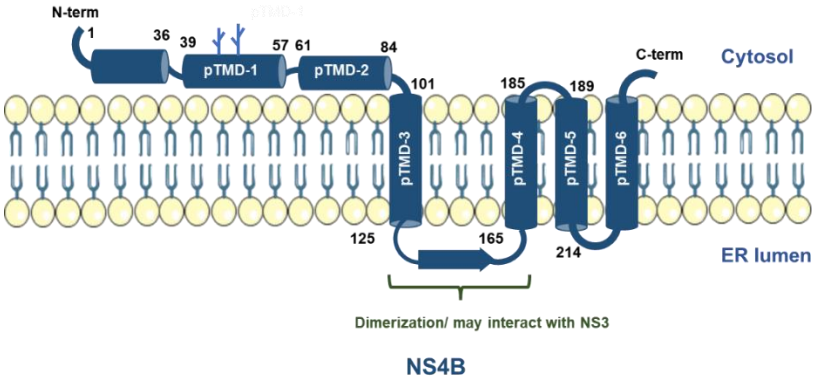


Figure 0-14 NS4B predicted transmembrane topology model with the functional marking of different regions

The exact transmembrane nature of NS4B is still a question (Figure 0-14); biophysical studies proved NS4B dimerization and oligomerization behaviour; a region from 125 to 165 amino acids showed dimerization character, and C-terminal 165 to 250 amino acids showed oligomerization behaviour, and there are two putative glycosylation sites mutations (**N58Q, N62Q)** showed these sites are essential for replication[126,127].

NS5 is the largest of all NS proteins with 103 kDa and processed by NS3 protease on the n-terminal side between NS4B and NS5 [81,91–93]. It is a multifunctional multi-domain protein like NS3. NS5 consists of an N-terminal methyltransferase (MTase) domain and a C-terminal RNA-dependent RNA polymerase (RdRp) domain (Figure 0-15).

Methyltransferase domain involved in cap synthesis**,** a three-step reaction involving NS3 and NS5 necessary for cap synthesis, first 5′-triphosphate of the mRNA is first converted to diphosphate by an RNA triphosphatase activity of NS3. The second reaction happens in two steps- the first step, the GMP intermediate, is formed from GTP via a phosphoramide bond, then it will be transferred to the diphosphate end of the RNA in the

second step by a guanylyltransferase to yield $G^{5\prime}$-$ppp^{5\prime}$-N. In a third reaction utilizing Sadenosyl-L-methionine (AdoMet) as the methyl donor, the transferred guanosine moiety is methylated by a (guanine-N7)-methyltransferase (N7MTase) to yield $^{7Me}G^{5\prime}$-$ppp^{5\prime}$-N (cap 0 structure) a second methylation reaction catalyzed by a (nucleoside-2′-O-) methyltransferase (2′OMTase) occurs on the first nucleotide three ′ to the triphosphate bridge to yield $^{7Me}G^{5\prime}$-$ppp^{5\prime}$-N_{Me} [128,129].

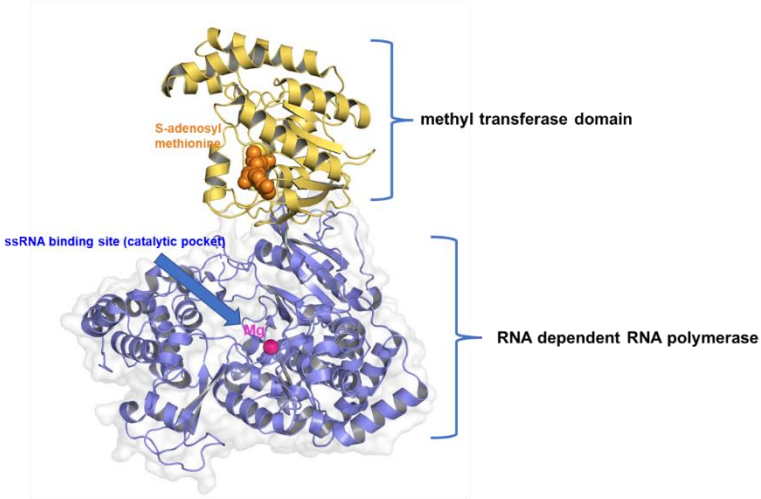


Figure 0-15 NS5 protein domain architecture (PDB ID- 5ZQK)

The methyltransferase domain ensures efficient de novo initiation and elongation steps by stimulating RNA template loading and increasing the binding affinity of NTPs to the priming site and the catalytic site[130]. In ZIKV, this methyltransferase domain of NS5 is essential for repressing IFN-β production, IRF3 activation, and RIG-I K63-linked polyubiquitination. Through these different mechanisms, NS5 may help DENV to surpass the dsRNA-stimulated IFN response [131].

RDRP domain carries out *de novo* RNA synthesis, has an internal nuclear localization signal, and exists as a dimer with both domains facing opposite directions[128,132,133]. RDRP domain adapts a canonical right-hand conformation with fingers, palm, thumb, and an

NLS (**316 to 415** with βNLS and α/βNLS sequences) containing region, which is essential for overall structural stability. The Finger domain (**273 to 315, 416 to 496, and 543 to 600)** base forms a concave surface that may accommodate the methyltransferase domain. The palm domain (497 to 542 and 601 to 705) contains four of six conserved motifs which are responsible for NTPs binding and catalysis, thumb domain (706 to 900) present at the C-terminal of RdRP with conserved motifs, with finger subdomains tips it will form a tunnel for RNA template, unlike other RDRPs in DENV. Other flaviviruses have a priming loop (initially Trp-795 and Thr-794, Ser-796 and Arg-737) which makes initial contacts with 3′dGTP to initiate denovo initiation and replication [134]. A linker region of NS5 (residues 266–271) involved in maintaining the interface interactions between methyltransferase and RdRP domains allows conformational freedom for both the domains[134]. NLS sequence helps NS5 to import into the nucleus, which may help NS5 from proteasome-mediated degradation in the cytosol [135,136].

Some putative sumo interacting motifs present in the N-terminal domain also appears to be important for viral replication - mutations in these motifs showed severe defects in replication of DENV[137]. NS5 also interacts with several host proteins like STAT2, URB4, and Impα2, which help to modulate the innate immune system and protect NS5 from degradation[138].

Complex and Dynamic Interactions of Dengue Virus Non-Structural Proteins

Flavivirus non-structural proteins interact with each other in a complex and dynamic manner. These interactions are essential for replication and polyprotein processing. NS4A, NS4B, NS2B, and NS2A being transmembrane proteins, show homomeric oligomerization, which may help in the replication-vesicle curvature and structure induction. NS4A and NS4B's are predicted transmembrane proteins and interact with each

other. Mutational analysis showed that abrogation of NS4A-NS4b interaction is determental to viral replication[115]. Recent studies showed NS2A might form an oligomeric complex near the pore opening of replication vesicle interacts with structural protein and, with the help of other nonstructural do the packaging viral genome into the capsid[139]. Along with these predicted transmembrane proteins, the dimer form of the NS1 protein with its wing domain loop spike interacts with replication vesicle from the lumen side of ER; this dimeric NS1 is essential for RC formation, and infectious WNV clones lacking NS1 gene were not able to form a functional RC[140]. NS4A N-terminal cytoplasmic loop showed to bend artificial membranes invitro, NS2B and NS2A with predicted viroporin topology may help maintain the environment inside the replication vesicle feasible for replication. Besides remodelling of ER membrane during the formation of IV and CMs and colocalization with dsRNA, how or whether the membrane-bound NS proteins play a direct role in RNA synthesis is unclear.

In the case of other heteromeric interactions, NS1 interaction with NS4A and NS4B was shown for flaviviruses through genetic assay, where a mutation in NS1 was suppressed by mutations in NS4A and NS4B[141,142]. Moreover, the fusion loop structure of NS1 wing domain induced curved membranes, Cytoplasmic loop of NS4A interacts with curved membranes, indicates their colocalization [111]. In DENV, NS1 protein G159-X-G161motif in wing domain interacts with NS4A-2K-NS4B precursor but not individual proteins, and this interaction is shown to be essential for replication with unknown mechanism and disturbing this interaction did not affect membranous vesicle induction and virus particle synthesis[121]. First of the predicted transmembrane region (40 to 76 amino acids) of NS4A and NS4 B's 84 to 146 amino acids region involved NS4A-NS4B interaction which is essential for replication of DENV, but this interaction may happen in NS4A-2K-NS4B

precursor which was already shown to interact with NS1 is essential for viral replication which may be necessary for the formation of replication complex[143]. The N-terminal part of NS2A interacts with NS3, and there is a 'basic' cluster comprising R22-K23-R24 in the N-terminal region of NS2A that is involved in virus assembly. The strand displacement function of NS3 along with NS2A is thought to be required for virus particle synthesis and for packaging the genomic RNA (Figure 0-16) into capsid[144]. Terminal 285 nucleotides of 3'-UTR interact with the cytosolic loop of NS2A, and this sequence may act as a signal for capsid packaging; during virion assembly, NS2A protein recruits viral RNA by its specific interaction with 3'-UTR and other structural proteins, NS3, and protease to the virion assembly site and helps in nucleocapsid formation[139].

DENV NS3 and NS5 interaction is the most critical viral component of the RC because both the proteins perform all the enzymatic activities required for RNA replication[61]. Functionally this interaction is a typical helicase-polymerase interaction which is seen in genome replication mechanisms in general. N-terminal (residues 320–368; known as bNLS) of NS5 interact with last 50 amino acids of NS3 C-terminal domain. This modulates NS3 helicase domain functions and stimulates NS3 helicase to unwind longer RNA duplexes which are required NS5 polymerase activity[145–147].

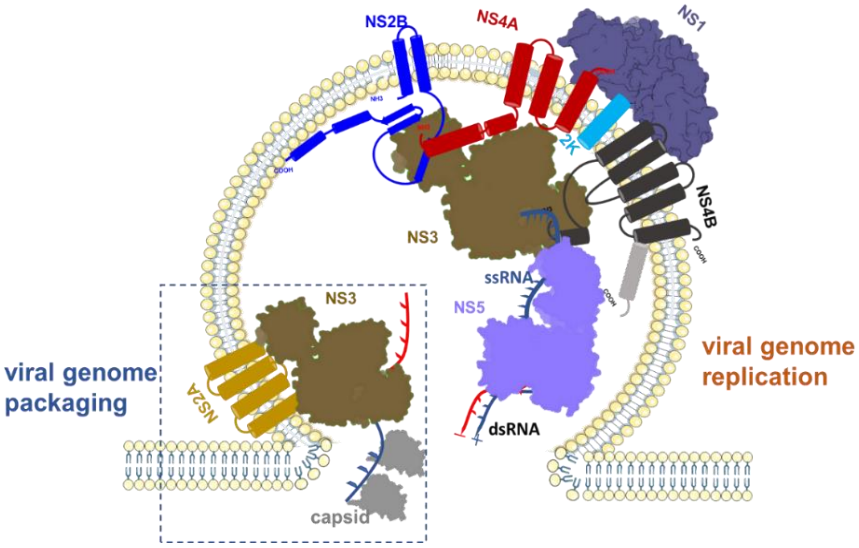


Figure 0-16 Schematic representation of viral genome replication and possible viral packaging sites

The helicase domain of NS3 interacts with the NS4A cytoplasmic C-terminal acidic EELPD/E motif, which appears to be functionally similar to the acidic EFDEMEE Motif of hepatitis C virus (HCV) NS4A, which is essential for regulating the ATPase activity of NS3hel and may act as a cofactor for helicase activity. The interaction of NS4A protein with NS3 will make NS3 sustain the helicase activity at low ATP concentration in WNV[140]. Yeast two-hybrid assay, coimmunoprecipitation, pulldown, and colocalization assays showed the interaction between NS4B and NS3 is required for dissociation of NS3 from single-stranded RNA, implying helicase activity modulation by NS4B[148,149]. Another NS-protein interaction, NS2B-NS3 interaction, is not only essential for the protease activity of NS3, but also brings specificity towards RNA as a substrate for unwinding in NS3. Among these nonstructural protein interactions, NS3-NS5 and NS2BNS3 interactions showed serotype specificity[132,150,151].

Some of these nonstructural proteins are also shown to interact with genomic RNA, which is essential for packaging and replication. NS2A, NS4A(weak), and NS3 were shown to

interact with the 3'-UTR, although NS2A showed strong interaction with terminal 283 nucleotides of the 3'-UTR necessary for replication and nucleocapsid synthesis. In the 5'UTR also, a structurally conserved SLA region interacts with NS5 and acts as a promoter for RdRP activity of NS5, both SLA and SLB structures in this region are essential for negative-sense RNA synthesis, and first, 12 bases of the SLA region form base-specific contact with NS3 helicase.

NS3's Multiple Roles and Interactions in Polyprotein Processing and Replication

As explained in the section preceeding this one, NS3 interacts with multiple nonstructural proteins (Figure 0-17) and has a role in multiple processes showing its importance as a centre molecule for viral multiplication inside the cell. How do these nonstructural protein interactions with NS3 impact its function, and their structure-function relationship will give the clues for drug design. Out of all those interactions NS3 made with other nonstructural proteins, interactions of NS2B, NS4B, and NS4A with NS3 are important to understand as these interactions affect NS3's protease and helicase activities.

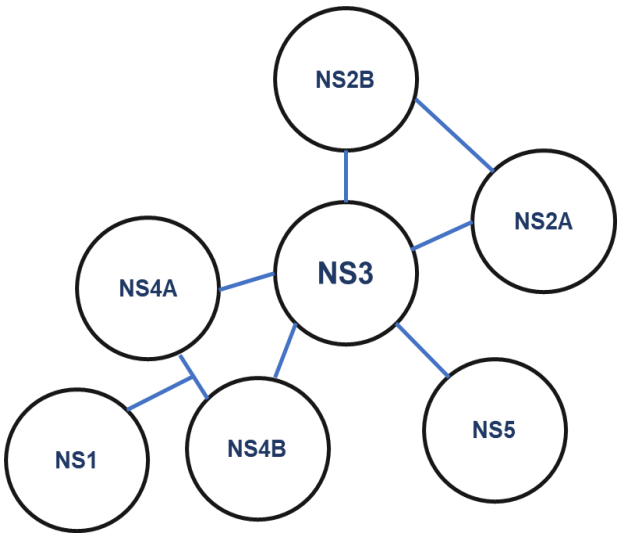


Figure 0-17 Schematic NS3 protein interaction with other nonstructural proteins

NS2B interacts with the N-terminal protease domain of NS3. In DENV, the protease is an independent structural domain of NS3 with a chymotrypsin-like fold and conserved serine protease motifs. The C-terminal half of NS2B (soluble cytoplasmic region of 45100 amino acids) acts as a cofactor for NS3 and forms a complex (NS2B3)[85,87,152], contributing one beta strand and a beta-hairpin to the chymotrypsin fold of the protease domain. More detailed description of the NS2b-NS3 protease domain interaction is presented in Chapter 2. A schematic depicting the complex interactions between NS3 and other NS proteins is shown in Figure 0-17.

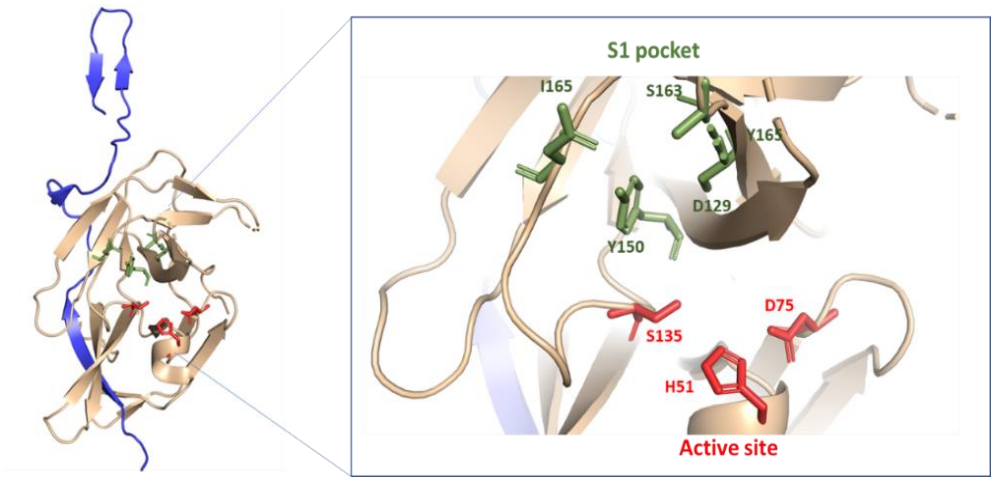


Figure 0-18 NS2B3 open conformation with the active site and S1 pocket residues highlighted (PDB ID-3L6P)

NS3 protease, like other serine proteases, has a catalytic triad "His50-Asp75-Ser136" arranged near the S1 pocket comprising ser163, Tyr165, Tyr150, Asp129, and the backbone residues at position 130 tyrosine and 132 asparagine helps to access the P1 basic residue by the active site in both WNV and DENV (Figure 0-18). S2 pocket, which interacts with additional basic residue and gives additional specificity to cleave only viral

polyprotein at specific sites, is still unclear; it was shown in WNV [153], but considering the second 'basic' residue fluctuates between P2 to P5, it needs further more clarification.

Structurally, compared with HCV protease, where the cofactor part of NS3 comes from NS4A and has some different secondary structures, the basic fold is precisely similar to NS3 of Dengue. However, these two proteins have different cleavage site recognition and cleavage specificity (HCV- C/S), and the P6 positional specificity of HCV protease having a negatively charged residue is essential for its cleavage specificity[154].

Like other naturally occurring proteases like trypsin and thrombin, NS3 also has an aspartate six residues before catalytic serine, which gives cleavage specificity after basic residue in trypsin and thrombin by holding side chain on this basic residue during peptide bond cleavage, but in NS3 mutation at this position has minimal effect on cleavage specificity of NS3[155]. Overall structural and mutational studies understanding giving the clues about the fold responsible for arranging the catalytic triad for cleavage of the peptide bond is the same, but other structural elements of the protein influence cleavage specificities at the P1 position. In trypsin and chymotrypsin, changing specific loops and flexible regions of the protein on the surface exchange their functional cleavage preferences after a particular residue. However, the residue held in the catalytic pocket, after which peptide bond cleavage happens, and how it comes into the catalytic pocket is not yet disclosed for most proteases[156,157]. Dengue protease has two different conformations, 'open' and 'closed,' but most of the substrate-bound crystal structures and solution NMR structures were in closed conformation with change in some part NS2B (77-87amino acids region) structural orientation towards the catalytic pocket. How these conformations interchanged and protease-activated and recognized specific cleavage sites between junctions, and the role of these conformations and residues in the initial

recognition of cleavage sites and the residues involved in S2 pocket formation are the questions to be answered.

Along with NS4B, NS4A also colocalizes with both NS3 and NS2B in the replication vesicle with dsRNA. NS4A interaction with NS4B and NS1 was essential for replication and virus particle synthesis. In ZIKV and WNV showed NS4A-NS3 interaction, whereas NS4A N-terminal cytoplasmic helix showed interaction with NS3 and modulated its ATPase activity [140,158]. However, this ATPase modulation was reported with a small stretch of NS4A sequence as an uncleaved protein. During polyprotein processing, a transient intermediate of uncleaved NS3-NS4A is required for the cleavage between NS4B-NS5[73,93,120]. After polyprotein processing, these cleaved NS3-NS4A will be processed into individual proteins and transported to ER membrane residing replication vesicles. After this cleavage, whether this interaction in the context of DENV replication is it shares the same interface or is there another dynamic interaction is possible or not is a question to be answered.

Scope of the Present Study

As evident from the review of the literature on DENV genomic RNA replication mechanisms and role of different non-structural proteins, their interactions with each other and with the viral RNA, several independent studies showed, using various experimental approaches, the criticality of NS protein interactions in viral replication. Of different NS protein interactions that are reported so far, interaction of NS3 with NS5, NS4b appears to be one of the critical interactions for DENV RNA replication. Modulation of NS3 helicase activity by NS3-NS4b interaction is observed in vitro. However, mechanistic details on how the interaction modulates the helicase activity is not known.

Similarly, NS2b interaction with NS3 is reported to be critical for the protease activity of NS3. Studies *in vito* (with purified recombinant NS3 protease and NS2b co-factor and DENV infection studies in cell lines) showed that NS2B is required for stabilization of the NS3 protease and thus protease activity, as well as for substrate specificity. This is despite the fact that the serine protease domain of NS3 has all the required elements for the protease activity. However, mechanistic details on how NS2B co-factor contributes (structurally) to the NS3 protease domain to make it active and brings about the substrate specificity is not known.

This study was focussed on biochemically characterizing the NS3-NS4b and NS3-NS2b interactions using recombinant proteins and enzyme assays for NS3 helicase activity and protease activity. The study was also focussed on detailed characterization of the conformation dynamics in NS3 protease domain and NS2b co-factor after interaction that might explain the mechanism of protease substrate specificity.

Thesis Work Objective

In vitro characterization of interactions between NS3 helicase domain and NS4b, and NS3 protease domain and NS2B co-factor proteins.

Specific Aims

- Structure-function characterization of NS3-NS4B complex and map interaction interface to deduce a mechanistic model of how NS4B modulates the function of NS3 helicase activity.
    - Express and purify NS4B and NS3 helicase domain in E.coli.
    - Biochemical and biophysical characterization of NS4B and NS3 proteins.
    - Biochemical characterization of interaction between NS3-NS4B.
    - Crystallization of NS3-NS4B complex for structural determination.

- Functional/enzyme assays for NS3 helicase activity to explain helicase activity modulation by the NS4B-NS3 interaction

- Structure-function characterization of NS2B-NS3 protease complex for its substrate specificity using in-silico, biophysical and biochemical techniques. o Express and purify NS2B-NS3 protease.
    - Develop a substrate peptide that would mimic the natural NS3 scission site and optimize a gel-based protease assay using the substrate and recombinant NS3-NS2b.
    - Crystallization of NS2B-NS3 complex along with the substrate for structure determination.
    - biochemical characterization of NS2B-NS3 protease substrate specificity.
- Find out the interaction between NS3 and NS4A map the interaction interface and the functional characterization of this interaction.

Chapter 1: Structural and Functional Characterization of NS3-NS4B Interaction

1.1 Introduction – Localization of NS3 and NS4B in RC and Scope for Interaction

As detailed in the Introduction chapter, in the DENV-infected cell, all the non-structural proteins, viral genomic RNA, and ds RNA intermediates of replication (Replication Complexes, RC) localize to specialized membrane-bound complexes on the ER membranes of various morphologies, collectively defined as replication organelles (RO)[45,50,159,160]. As the viral polyprotein is synthesized, stretches of the polyprotein corresponding to NS2A, 2B, 4A, 4B, and a small peptide, 2K, connecting 4A and 4B (transmembrane proteins) are inserted into the ER membrane. NS3 and NS5 proteins (soluble proteins), on the other hand, reside in the RO lumen, whose contents are in a continuum with the cytoplasm and are the major enzymes involved in replication (Figure 1-1). However, NS3 is anchored to the membrane surface through interactions with NS2B, NS4B, NS4A, and NS2A. NS4B is a predicted transmembrane protein without any sequence similarity with other known proteins and does not have any conserved motifs but essential for replication, as mutations in NS4B result in defects in replication and packaging [127,161,162].

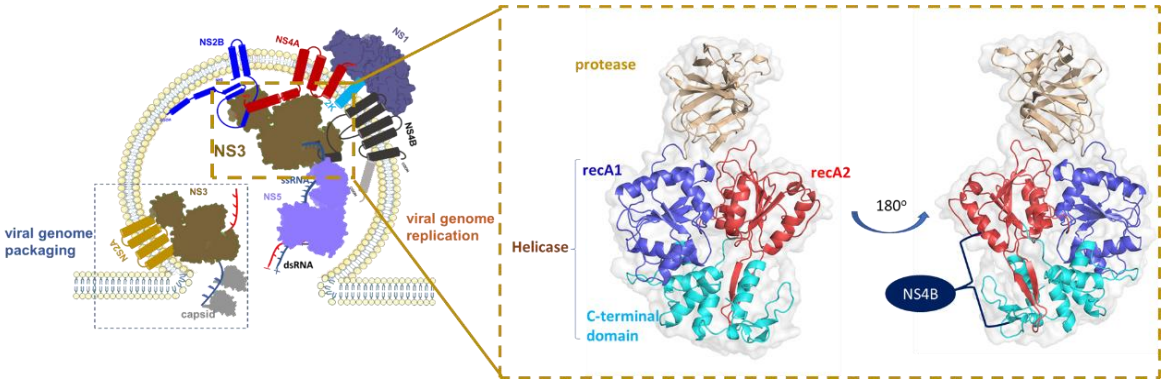


Figure 1-1 Replication organelle with a zoomed-out image of NS3 structure cartoon representation with domain architecture and proposed NS4B interaction site on NS3

Briefly, structural features of DENV NS3 are described here. NS3 is a multidomain multifunction protein with N-terminal one-third part of the 619 amino acid long protein (in DENV serotype 1) folds into an independent domain with a canonical chymotrypsin-like fold forming a serine protease. This domain, along with the part of NS2B proteins cytoplasmic region, acts as a co-factor and is involved in the proteolytic processing of the polyprotein. The C-terminal reminder of the protein folds into three distinct structural domains: two recA domains and an alpha-helical c-terminal subdomain (present in most of the monomeric SF2 helicases) and together functions as an RNA helicase. A flexible linker connects the N-terminal protease domain to the helicase domain. The crystal structure of the helicase domain alone[95] or full-length DENV NS3[163,164] for serotype-2 and serotype-3 are published. The subdomains recA-1 and recA-2 have the canonical sequence motifs (I, Ia, II, III, IV, V, VI) of a P-loop containing nucleoside triphosphatase (ATPase activity) and SF2 family RNA helicase. The C-terminal subdomain (CTD) is seen in all flavivirus NS3 proteins. Based on the structural features and canonical helicase sequence motifs, NS3 is categorized into the DExH subfamily RNA helicases of the SF2 superfamily of monomeric helicases[107]. This family of helicases loads onto a free 3'-end of the polynucleotide and unwinds the duplex in the 3'- to 5'- direction.

Different ligand-bound structures of DENV NS3 helicase[107] show that ssRNA binds in the cleft formed in the recA domains and the CTD interface, whereas an ATP molecule binds in the interface of the two recA domains. A structural mechanism of ATP hydrolysis-coupled RNA unwinding is proposed. The RNA binding and ATP hydrolysis lead to the inward movement of the P-loop (motif Ia), a twisting motion of the recA domains on each other, and the outward movement of CTD away from the recA domains.

These domain motions, especially recA domain motions, lead to the translocation of the RNA in the RNA-binding cleft coupled to every cycle of ATP hydrolysis. However, several aspects of the helicase mechanism are not understood completely. For example, the role of the CTD subdomain in the helicase mechanism is not understood well. The CTD is involved in interactions with other NS proteins, NS4B [148,161] and NS5[147]. In other SF2 superfamily members wherein subdomains analogous to the CTD are present (domains additional to recA-like domains), those subdomains are used for interactions with other proteins that act as regulators of the helicase function[165,166]. Several experimental pieces of evidence support the idea that NS3–NS4B interaction is critical for viral replication. Interestingly, DENV NS3 helicase, other flavivirus helicases, and a few other superfamily members show poor RNA helicase activity without interacting proteins[165,166]. The in vitro helicase activity of DENV NS3 helicase is enhanced upon interaction with NS4B protein[149]. However, the mechanism of the helicase activity modulation is not known.

NS4B is a predicted transmembrane protein and co-localizes along with NS3 in replication complexes on ER membrane[45]. The previous studies[167] provided the first biochemical evidence for two helices in the N-terminus of NS4B that peripherally associate with the membrane (named TM1 and TM2) and three transmembrane helices [spanning regions 101- 129(TM3), 165-190(TM4), and 217-244(TM5) in DENV serotype-2 NS4B sequence] in the C-terminal half of the protein. As per this proposed membrane topology, the N-terminal (until the beginning of the TM3) region of the protein would be on the ER lumen side, and the loop connecting the TM3 and the TM4 (cytosolic loop) is the only possible interaction site with the NS3. Studies for fine mapping this interaction based on this topology map, specifically between the NS4B cytosolic loop and

NS3. However, did not rule out the possibility that other regions of NS4B also interact with NS3[45,148]. Contrary to current knowledge, a recent study showed that the N-terminal region of NS4B, but not the cytosolic loop, is an important determinant of interaction with NS3[168]. To add support to this, a previous mutation-based genetic analysis of NS3NS4B interaction showed a lethal point mutation in the NS4B N-terminal region L28A has a suppressor mutation in CTD of NS3 Y601D[161].

Interestingly, the N-terminal half of NS4B (spanning the TM1 and TM2 helices) plays a role in modulating the innate immune response by inhibiting IFNα/β signalling, probably through direct interaction with STAT1, and modulating its nuclear localization [169]. Since STAT1 is a cytosolic protein, it is reasonable to think that the N-terminal 100 amino acids of NS4B are localized on the cytosolic side of the RO in contrast to the proposed topology for NS4B. Thus, the N-terminal region of NS4B may be an interacting site for NS3. For this study, the idea that the N-terminal 100 amino acids region of NS4B is on the cytosolic side of the RO is favoured. The Results and Discussion sections present other reasons for making this assumption.

As NS3 and NS4B proteins and their interaction are critical for viral replication, they are potential drug targets[11,146,162,170,171]. Recently, a small-molecule drug, JNJ-A07, that blocks NS3–NS4B interaction is shown to be a highly effective pan-serotype DENV inhibitor[11]. A complete understanding of the NS4B–NS3 interaction interface and the mechanism by which NS4B can modulate the helicase enzyme activity of NS3 will significantly help in devising such drug strategies. Our study thoroughly characterized the NS4B–NS3 interaction and provided a map of the NS4B–NS3 interaction interface. Our studies were also aimed at understanding the structural mechanism of the helicase activity enhancement upon NS3 interaction with NS4B. Upon interaction with NS4B,

there are significant tertiary structure changes in NS3. Based on these observations, a plausible mechanism proposed for the NS4B interaction-dependent enhancement of NS3 helicase activity wherein the interaction between the proteins modulates the dynamics of the CTD domain motion towards and away from the RNA-binding cleft, leading to helicase activity enhancement.

1.2 Material and Methods

1.2.1 Dengue Virus sequences and Sequence Predictions

The sequences of the non-structural proteins of the dengue virus used in this study are from NCBI GenBank accession number JN903579. Transmembrane helix predictions on the sequences are made with TMHMM, Phobius, or PSIPRED MEMSAT-SVM by submitting the NS4B sequence on the respective servers.

To check if there are any conserved secondary structures in the 3′-UTRs of flavivirus genomes, first analyzed, the last 120 nucleotide sequences of the DENV serotype-1 genome (corresponding to the domain III region of the 3′-UTR). And then performed a multiple sequence alignment of this sequence with other flaviviruses' corresponding 3′UTR sequences.

For predicting the secondary structures in the DENV1 3′-UTR sequence, an RNA structure application on the web server was used (https://rna.urmc.rochester.edu/RNAstructure.html). RNA multiple secondary structure alignment with 14 different flavivirus sequences (including DENV serotypes 1, 2, 3, and 4 3′-UTR sequences) on the RNAalifold server (http://rna.tbi.univie.ac.at/). The RNA and DNA stem-loop forming oligos that are used in the helicase assays (see below) were

designed based on the secondary structure prediction in the RNA structure web application.

1.2.2 Bacterial Two-Hybrid (BACTH) Assay

A bacterial two-hybrid assay (BACTH) was performed following the protocol described earlier[172]. Full helicase domain (residues 178–619 as per DENV serotype-1 sequence numbering scheme) or only the recA domains (residues 178–478) or recA-2 and CTD subdomains (residues 313–619) or only the CTD subdomain of NS3 helicase (residues 478–619) were cloned as a C-terminal fusion with the T25 fragment of the adenylate cyclase of the BACTH assay system, in the pKT25 plasmid. The resulting plasmids are named pSM61, pSM63, pSM64, and pSM62, respectively. Similarly, coding DNA of full-length NS4B1-249 (pSM55) or N-terminal 57 residues (plasmid pSM59) or NS4B residues 58–249 (plasmid pSM60) from DENV serotype-1 is cloned as an N-terminal fusion of the T18 fragment of the adenylate cyclase in the pUT18 vector. To perform the BACTH assay, different combinations of pKT25-NS3 helicase and pUT18-NS4B plasmids were used for transforming the E. coli BTH101 strain. The transformant cells were plated on an LB agar plate containing 50 µg/mL kanamycin and 100 µg/mL ampicillin. The transformants that grew on the plate were patch streaked on a MacConkey-maltose agar indicator plate supplemented with 50 µg/mL kanamycin, 100 µg/mL ampicillin, and 0.5 mM IPTG and incubated at 30°C for 96 h. The plasmids pUT18-Zip and pKT25-Zip, containing the interacting domains of the GCN4 leucine zipper, were used as a positive control in the assay. Empty pUT18 and pKT25 plasmids were used as a negative control for the assay. Each experiment was done in triplicates.

1.2.3 Homology Structure Modeling of NS3, Ab-Initio Structure Prediction of NS4B

A homology model of DENV serotype-1 NS3 was built using the DENV serotype-3 NS3

structure (PDB ID: 2WHX) using the 'SWISS-MODEL' server (https://swissmodel.expasy.org).

Two newly released machine learning algorithms based on protein-structure-prediction software, AlphaFold2.0[173], and RoseTTAfold[174], are used to predict a structure model of DENV serotype 1 NS4B. AlphaFold2.0 predictions used the AlphaFold Colab notebook (Shared publicly through a Creative Commons At-tribution-NonCommercial 4.0 International license). This AlphaFold Colab notebook uses the BFD database for prediction after a multiple-sequence alignment. An energy-minimized predicted model PDB file was downloaded along with the pLDDT scores. Structure Figures were prepared with PyMOL. The DENV serotype-1 NS4B sequence was also submitted on the RoseTTAfold server for structure prediction. The five best structure models, as scored based on angstrom error estimate per residue parameter, were used. Since there was no significant difference in the predictions using AlphaFold2.0 or RoseTTAfold, finally the structure predicted by AlphaFold2.0 was used for further analysis.

1.2.4 Crystallization of NS3 Helicase - NS4BN57 Complex

NS3 protein concentrated to 5mg/ml in 20mM Tris, pH7.5 buffer containing 150mM NaCl, 5mM DTT, 10% glycerol buffer. Crystallization conditions were tried for NS3, NS3:NS4B(1:3), and NS3:NS4BN57(1:3) protein in all ratios NS3 concentration was kept at 3.5-5mg/ml. Crystallization drops were set in a 1:1 ratio (1ul protein and 1ul of crystallization buffer) with 0.1M MES pH 6.0 and pH6.5 8-15% PEG4000 conditions in a hanging drop mode.

1.2.5 Molecular Docking

For performing molecular docking simulations, an energy-minimized structure of the homology model that built for DENV serotype-1 NS3 and the predicted structure of

NS4B. A molecular docking simulation was performed on the Haddock server[175]. Initially a rigid body; global, blind docking run on the ClusPro[176] server (https://cluspro.bu.edu) to get a possible docking pose and information about the interacting residues. Using the set of residues in the interface of the docked pose from the clusPro docking run, a local, flexible simulated docking was set up in a Haddock 2.4 (https://wenmr.science.uu.nl/haddock2.4) docking server. The N-terminal 57 residues of NS4B were set as active and flexible regions in the docking run. The docked pose with a rank '1' docking score was used for further analysis.

1.2.6 Molecular Dynamic Simulations

Molecular dynamic simulations were performed on the NS4B-docked NS3 structure using GROMACS software. Two different simulation runs were set up, one with the NS3 structure alone and another with the NS3–NS4B complex structure (correspond-ing to the best docking pose from the molecular docking studies). Both proteins were solvated with water (water model-TIP3P), and no ions were added. The NS4B-docked NS3 protein model was enclosed in a box of 14.903 nm, 14.910 nm, and 14.906 nm dimensions. The system was energy-minimized in an NVT ensemble and equilibrated for 125 picoseconds in the NPT ensemble. Each system was simulated using Gromacs-5.1 with a CHARMM36 forcefield with a 1fs step time. Periodic boundary conditions were applied in the simulation. The system was subjected to simulation production for 20ns at 303 K after equilibration.

1.2.7 Bacterial Expression and the Purification of NS3 and NS4B Proteins

The helicase domain of DENV NS3 (NS3Hel) is expressed in a Rosetta DE3 (E. colibased) expression system and purified the protein using six-histidine-tag affinity chromatography, following a protocol previously published[95] with a few modifications.

DNA clones with NS3 and NS4B coding regions, pDV1-419NS2B3 and pDV1419NSP4b, respectively, were a donation from Dr. E. Sreekumar, Rajiv Gandhi Centre for Biotechnology, Thiruvananthapuram. The coding DNA corresponding to NS3 residues 178 to 619 was PCR-amplified (primers SM31 and SM6) cloned into pET24b vector using Nhe I and Xho I sites so that the resultant protein has a 6-histidine-purification tag as a C-terminal fusion. The coding region of DNA corresponding to NS4B residues 1–249 was PCR-amplified (primers SM96 and SM10), respectively, and cloned into pET24b with a C-terminal six-histidine tag. All clones were sequence verified.

For expression, competent cells of Rosetta (DE3) E. coli were transformed with pSM10 plasmid. Transformant cultures were grown in LB Broth supplemented with 50 μg/mL kanamycin and 34 μg/mL chloramphenicol to an optical density at 600nm (OD600) of 0.6. Protein expression was induced by 0.3mM isopropyl β-D-1thiogalactopyranoside (IPTG), after which cultures were grown at 14°C for 20 hours. Cells were harvested by centrifugation and lysed by sonication in 20mM Tris-Cl, pH 7.4, buffer containing 1M NaCl. Protein purification was done from the clarified ly-sates using 6-His-tag affinity to a Ni-NTA agarose column. The protein was bound to Ni-NTA with 20mM imidazole containing buffer, followed by a 50mM imidazole wash step and elution with 300mM imidazole-containing buffer. As a final purification step, the 300mM imidazole elution fraction from the His-tag-affinity purification step was concentrated and loaded on to a Superdex-75 Increase 10/300 GL gel-filtration column. Elution was done with 20mM Tris-Cl, pH 7.4, buffer containing 150mM NaCl and 1mM EDTA. The purity of the protein was assessed by SDS-PAGE analysis of the concentrated fraction of the gel-

filtration chromatography eluent. Protein concentration was estimated by recording the absorbance at 280 nm and using the E1% value of 13.54.

For the expression and purification of the NS4B full-length, previously published bacterial expression and protein detergent extraction protocol was adapted[127]. The plasmid pSM6 was used to transform BL21(DE3) E. coli. The transformant bacterial culture was grown for 16 hrs at 18°C after protein expression induction with 0.5mM IPTG. The cells were harvested by centrifugation and lysed by sonication in 20mM TriCl, pH 7.4, containing 150mM NaCl (TN). The NS4B was in the pellet fraction after the clarification step of centrifugation at 30,000× g for 30 min. The protein was solubilized and extracted out of the pellet by detergent extraction protocol using incubation in TN buffer containing 3% LDAO and 3M guanidine hydrochloride for 1 hour at RT. After a centrifugation clarification step (30,000× g for 30 min), the clear extracts were diluted to reduce the concentration of the detergent and Gd.HCl to 1% and 1M, respectively, before loading onto a Ni-NTA agarose column. The protein was eluted as a pure fraction with 200 mM imidazole and 0.1% LDAO-containing TN buffer. The fractions containing the pure NS4B protein were concentrated, as assessed by SDS-PAGE analysis. The concentrated fraction was loaded on to a Superdex-75 increase 10/300 GL gel filtration column and eluted with TN buffer containing 0.05% LDAO and 1mM EDTA. NS4B protein eluted at different elution vol-umes in the gel-filtration chromatography. Each elution fraction was collected separately and concentrated by ultra-filtration. Protein concentration was estimated by recording Abs.280 nm and using an E1% value of 12.6.

NS4B also purified without using a detergent extraction protocol. The protein expression conditions differed from the above-described protocol in that the incubation temperature was set to 37°C, and protein expression induction was done by adding 1mM IPTG to the

culture for 4 hours. The cell lysis was done by sonication, and lysates were clarified using centrifugation. The pellet fraction after the clarification step was incubated with 6M Gd.HCl containing TN buffer overnight. After a clarification step by centrifugation, the clear fractions from the denaturant extraction step were loaded onto a Ni-NTA agarose column. The protein was slowly refolded on the column by changing the concentration of Gd.HCl in the wash and elution steps sequentially to 3M and 1M. The protein thus eluted off the column with 300mM imidazole and 1M Gd.HCl containing TN buffer was further diluted so that the concentration of the denaturant was reduced to 0.05M. The protein was kept on ice at this step. Subsequent purification steps included the protein concentration using ultrafiltration and then gel filtration chromatography on a Superdex75 column. The protein thus obtained is more than 95% pure.

The N-terminal 57 residues region of NS4B (NS4BN57) was expressed in a similar way. The protein was expressed from a pET24b vector in BL21(DE3) E. coli cells. The expression conditions were: culture growth temperature set to 30°C and protein expression induction with 0.5mM IPTG. The NS4BN57 is expressed in the form of a soluble protein. The protein was purified from the clarified lysate (in TN buffer) using Ni-NTA His-tag affinity chromatography.

1.2.8 Far-UV Circular Dichroism (CD) and Intrinsic Tryptophan Fluorescence Spectroscopy

Far-UV CD measurements were done using a Jasco 1500 spectropolarimeter. The protein concentration was 0.2 mg/mL in each case. Spectra were recorded at 50 nm/min scan speed in the wavelength range of 200–250 nm. Bandwidth was set to 1 nm. Each spectrum was an accumulation of three scans. Each spectrum was blank-corrected (TN buffer alone or TN buffer along with 0.05% LDAO when NS4B purified by detergent extraction method was used) and smoothened. Mean residue ellipticity, θm.r.e, was calculated using

the equation [θ]m.r.e = 100×θ/(CNl), where θ is the observed ellipticity at a given concentration of protein, C is the molar concentration of the protein, N is the number of residues in the protein, and l is the cell path-length in centimetres.

For studying secondary structural changes in NS3Hel or NS4B or both upon interaction, the method described by Greenfield NJ, 2015 was followed[177]. The far-UV CD spectra were recorded for NS3Hel alone or NS3Hel: NS4B protein mixtures at different molar stoichiometries (1:1, 1:2, 1:3). The NS4B protein that is used in this assay was purified following the detergent-free protocol described above. The data was analyzed by converting the ellipticity values to θm.r.e. In the mixture of the two proteins, the θm.r.e was calculated using the cumulative molecular weight of the monomer mass of NS3 and the monomer mass of NS4B (51.8 kDa for NS3Hel + 27.8 kDa for NS4B). If there were any secondary structure changes, upon interaction, it was expected that the observed θm.r.e for the mixture would be different from the expected value.

Intrinsic tryptophan fluorescence spectra were recorded using an FLS 1000 fluorimeter (Edinburgh Instruments, UK). The protein was used at a 0.05 mg/mL concentration for recording the fluorescence spectra. The excitation wavelength was set to 295 nm, and the emission wavelength range was set to 300–500 nm. The excitation and emission bandwidths are set to 1 nm. Spectra were blank-corrected and smoothened using the Fluoracle (Edinburgh Instruments, UK) software. For each spectrum recorded on the NS3:NS4B mixture, the corresponding NS4B concentration spectrum was subtracted, following the protocol described earlier[178].

1.2.9 Bio-Layer Interferometry (BLI)

A Bio-layer interferometry assay was performed to test the interaction between NS3Hel and NS4B proteins. Measurements were recorded using the Octet (FortéBio) platform using High Precision Streptavidin (SAX) Biosensors (FortéBio) (courtesy of FortéBio). Using standard reagents and protocols supplied by the manufacturer (FortéBio), the NS4B protein was biotinylated. The NS4B protein was loaded onto the sensor at 2 μg/mL concentration. After a wash step with BLI buffer, the NS4B-bound sensor was incubated with different concentrations of NS3Hel, (20, 6.647, 0.7407 μM) in TNE with 0.05% tween-20. Binding was followed for 200 secs, where it reached equilibrium. The biosensor was then incubated in the buffer for 200 secs to measure the dissociation reaction. The shift in the wavelength of light reflecting from the sensor (in nm) is measured in real-time during the binding and dissociation phases. The binding response curves (sensorgrams) for the indicated NS3 helicase protein concentrations over the NS4B-immobilized sensor are plotted. The equilibrium dissociation constant (KD) of binding was estimated from the curve-fitting on the response curves using Octet software (FortéBio).

1.2.10 Liposome Preparation and Co-Floatation Experiments

Small unilamellar vesicles (SUVs) were prepared, as described by us earlier, with synthetic purified lipids POPC and cholesterol in a 1:1 ratio. A lipo-some co-flotation assay was performed using NS4B purified using a detergent extraction protocol (200 μg) and 0.5mM of lipid vesicles in a final volume of 200μL TNE buffer. Briefly, the liposome–NS4B protein mixture was placed above a 40% sucrose bed (4 mL), which was topped by two layers of 20% and 5% (4 mL each) sucrose. After ultracentrifugation for three h at 200,000× g, fractions from the top (4 mL) and bottom (4 mL) were collected. Aliquots

from each assay's top and bottom fractions were analyzed with Western blotting using anti-His-tag monoclonal antibodies.

1.2.11 Electrophoretic Mobility Shift Assays for Helicase Activity Measurement

An adapted protocol described by Xu et al. (2005)[95] was used for helicase activity measurement using an electrophoretic mobility shift assay (EMSA). The protocol was further optimized with an RNA and a DNA stem-loop oligos that were designed (see dengue virus sequences and sequence predictions section above) to mimic the 3′-SL of the DENV. Briefly, an RNA oligo with the sequence 5′UCUACAGCAUCAUUCCAGGCACAGAACGCCAAAAAAUGGAAUGGUGCU GUUGAAUCAACAGGUUCUUUUU-3′ with FAM (6-Carboxyfluorescein) fluorophore attached at the 5′- end of the oligo was procured from GenScript, USA. The RNA oligo was visualized after the EMSA gel run using the FAM fluorescence. A 66-nucleotide DNA oligo was synthesized with sequence 5′-TCTACAGCATCATTCCAGGCACAGAACGCCAAAAAATGGAATGGTGCTGTT GAATCAACAGGTTCT-3′. Before using the oligonucleotides in the EMSA assays, a 94° C denaturation and slow annealing step was included to ensure a proper stem-loop structure. The helicase assay was done by mixing the RNA (0.75 picomoles per reaction) or the DNA oligo (3 picomoles per reaction) with 20 and 50 molar excess of NS3Hel protein alone, respectively, or with NS3Hel: NS4B in different (1:1, 1:2, 1:3) molar ratios. The final reaction volume was adjusted to 10 µL to achieve a final buffer concentration of 20 mM Tris-Cl, pH 7.4, containing 60 mM NaCl and 1mM MgCl2. The assay was started by adding ATP to the reaction mixture at 2mM concentration for RNA oligo and 4mM for the DNA oligo reaction mixture. The reaction was stopped by adding the EMSA loading dye (containing 5% glycerol and 10mM EDTA) to the reaction mixture at the

indicated time points. EMSA was performed on 15% native Tris Borate EDTA (TBE)polyacrylamide gel. Before the experiment, it was first confirmed that the 3′-SL DNA/RNA oligos actually form a stem-loop structure by analyzing the oligos on EMSA along with the heat-denatured sample of the oligo in the next lane. The mobility of the heat-denatured oligo in EMSA matched that of unwound stem-loop/duplex substrates used for the DENV NS3 helicase by others [20]. The oligo samples that were not denatured showed slower mobility in EMSA compared to denatured/unwound oligo, confirming that the DNA/RNA oligos formed stem-loop structures as predicted. The RNA or the DNA oligo alone in helicase assay buffer served as negative controls for the respective experiments. NS4B alone and NS3 without adding ATP along with the RNA and the DNA oligo are also included as controls to check that the unwinding activity is not because of some unknown factors coming from the protein preparations or buffer components. After the electrophoretic run, staining for DNA was done using SYBR Gold as per the manufacturer's protocol, and the gel images were recorded. For experiments with RNA oligos, the gels were imaged directly after the electrophoretic run using the Alexa 488 filter of a gel imager. The helicase activity was estimated by taking pixel intensity (from the images of the stained EMSA gels) of the ssRNA (un-wound RNA stem-loop) or ssDNA band over the cumulative intensity of the single-stranded and the stem-loop bands. For comparing helicase activity in the absence and in the presence of NS4B, normalized the helicase activity in each case to helicase activity with NS3 alone. The % helicase activity in each case is plotted in a bar plot along with + standard error. The average helicase activity with NS3Hel alone at one hour past ATP addition to the reaction mixture was taken as 100%.

NS3Hel RNA duplex unwinding activity was also assessed using a molecular beacon assay published earlier [42]. Two RNA oligos: CY5- 5′-GACGUCAGUUGUUAGUCU ACGUC -3′—BHQ2 (wherein CY5 is a fluorophore with excitation and emission maximum at 630 and 670 nm, respectively, and BHQ2 is a black hole quencher) and 5′- AGACUAACAACUGACGUC UUUUUUUUUUUUUUUUUUU-3′, with complementary sequences (underlined text in both sequences), are used to form an RNA duplex. The helicase reaction mixture contained 30 nM of dsRNA with 100 nM of NS3Hel protein alone or NS3Hel: NS4B/NS4BN57 in a 1:3 molar ratio. The final reaction volume was adjusted to 25 µL to achieve a final buffer concentration of 20 mM Tris-Cl, pH 7.4, containing 15 mM NaCl and 2mM MgCl2. The assay was started by adding ATP to the reaction mixture at 2mM concentration, and the fluorescence intensity of CY5 (at 670 nm) was recorded for the next 30 min. Fluorescence intensity (FAU at 670 nm) before the addition of ATP was taken as the starting fluorescence value (F0), and fluorescence intensity at 30 min after starting the assay as the assay end point value (F30). As the fluorescence of CY5 would be quenched by the black hole quencher after the duplex is unwound and the labeled oligo formed a stem-loop structure, the F30 value would be lower than F0. The F0 – F30 was taken as 100% helicase activity of NS3Hel. The percent increase in helicase activity in the presence of NS4B or NS4BN57 was calculated using the formula (F'0- F'30) X 100/(F0- F30), where F'30 and F'0 are endpoints and initial fluorescence values in reaction mixtures with NS3:NS4B or NS4BN57 in a 1:3 molar ratio.

1.2.12 ATPase Assay

The ATPase activity of NS3Hel was measured by quantifying the release of free phosphate (Pi) following ATP hydrolysis by malachite green assay in the presence of a DNA stem-loop following a protocol described earlier [43]. The assay was performed in a final volume of 30 µL by mixing NS3Hel (50 nM) protein alone or NS3Hel: NS4B/NS4BN57 in a 1:3 molar ratio. The NS3 protein was pre-incubated with 0, 0.3, 0.6, 1.2, 2.4, and 4.8 µM of the DNA stem-loop oligo in a reaction buffer of 20 mM TrisCl, pH 7.4, 1 mM MgCl2, and 60 mM NaCl. The ATPase reaction was started by adding 1mM of ATP and incubated further for 40 min at 30 °C. The reaction was stopped by adding 20 mM EDTA, and aliquots were collected every 10 min and stored at 4°C until further processing. A total of 10 µL malachite green reagent (Sigma, USA) was added to a 40 µL sample and incubated at room temperature for 30 min to form a complex with molybdate and free orthophosphate. Samples were transferred into a 96-well plate, and the absorbance at 620 nm was recorded using a Thermo Varioskan microplate reader. From the Abs.620nm values, the orthophosphate concentration, and thus the Pi released from the ATPase activity of the helicase, was estimated using a calibration curve with known concentrations of orthophosphate.

For Michaelis–Menten kinetics studies on NS3Hel, the ATPase assay was done with different concentrations of the substrate (stem-loop DNA oligo), keeping the ATP concentration and enzyme concentration in each experiment constant. Initial velocity measurements at different substrate concentrations and fitting the data into an MM-kinetics model was done using the ICEKAT web application (http://icekat.herokuapp.com/icekat accessed on 12th November 2021). Slopes from the linear range of each malachite green absorbance versus time kinetic trace are taken as the

initial rates. The kinetic data is fit to an MM-kinetic model using a non-linear regression method either in the ICEKAT application or in Graphpad Prism software.

1.3 Results

1.3.1 N-Terminus 57 Residues of NS4B Are Enough to Interact With NS3

To fine-map the interaction regions on NS3 and NS4B, a bacterial adenylate cyclase based two-hybrid (BACTH) assay were performed. The interaction of full-length NS4B, only the N-terminal 57 residues region and N-terminal 57 residues truncated NS4B proteins with different subdomains of the NS3 helicase were tested in the BACTH assay. The rationale for selecting different truncations of NS4B for testing the interaction is as follows: earlier sequence predictions and biochemical studies[167] showed that most of the protein inserts into the ER membrane (TM3, TM4, and TM5) and only the N-terminus 100 residues (or the loop connecting TM3 and TM4, if the protein inserts into the membrane in orientation as proposed by Miller et al.[167]) might be accessible for NS3 interaction. Hence, the focus and interest are to see if the N-terminal region of NS4B alone can interact with NS3. Even in the N-terminal 100 residues, two membrane associating helices are predicted, although there is an inconsistency in previous transmembrane region predictions reported in different studies. This experiment included only the N-terminus 57 residues (till the end of the predicted first helix) as this region may be the most distal structural element of the NS4B protruding into the cytosol side of the RO and thus most accessible to NS3 for interaction.

The interaction between the proteins was assessed qualitatively by observing red color development on the MacConkey/maltose agar indicator plates surrounding the double transformed bacterial colonies and comparing it with colour development on the control plates. As expected, and confirming earlier observations by others[148,161], the full-length

NS3 helicase and full-length NS4B transformed colonies showed red color development, nearly comparable to that of the positive control bacterial patch (Figure 1-2). Furthermore, it is also observed that recA1 and recA2 subdomains alone of the NS3 helicase, as well as recA2 and CTD and the CTD subdomain alone, interacted with NS4B comparably to the full-length NS3 helicase (Figure 1-2). This implies that the NS3–NS4B interaction interface may spread over a large surface area of NS3, spanning different subdomains of the helicase.

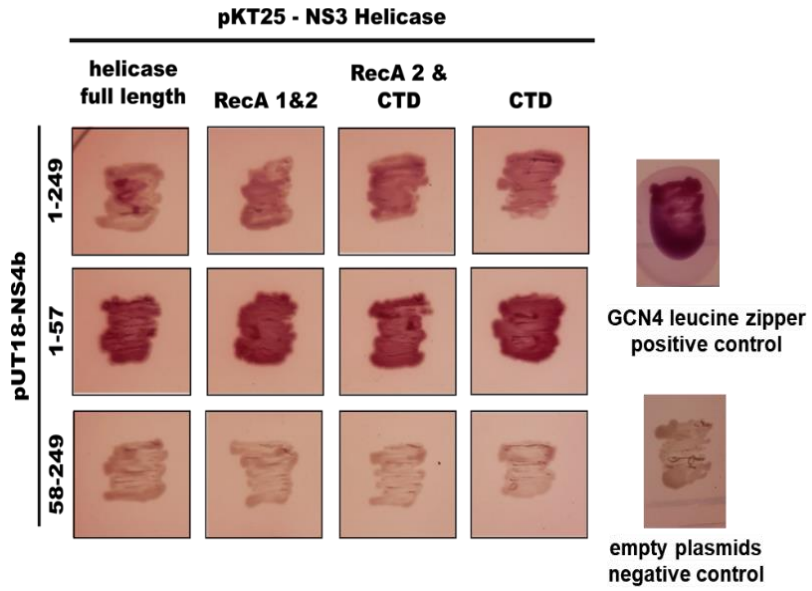


Figure 1-2 BACTH assay of NS3-NS4B interaction -Images of colony patches on MacConkey/maltose agar indicator plates with different truncations of NS3 and NS4B constructs

Surprisingly, when only the N-terminus 57 residues region of the NS4B was tested in the assay, the color development and colony growth were much more robust than those observed with full-length protein constructs (Figure 1-2, middle row), indicating a more robust interaction strength. Stronger interaction between the N-terminal 57 residues fragment of NS4B with the NS3 helicase full-length protein or with different subdomain variants of it in different replicates of the experiment. On the other hand, the N-

terminal57-residues deleted NS4B construct did not show any interaction with NS3 (Figure 1-2, bottom row, no color development). These observations align with our thinking that the N-terminal end of the NS4B is the only region accessible for NS3 interaction. Supporting this idea, a recent study[168] also noted that the N-terminal region of NS4B (residues 51– 83) is enough to interact with NS3. Thus, from our BACTH assays, it is apparent that the

N-terminal 57 residues of NS4B (or a sub-region within) is enough for interaction with NS3, and the interaction interface on NS3 spans all three subdomains of the helicase domain. However, from these experiments, it cannot exclude the possibility of other regions, especially the long loops connecting the predicted transmembrane helices, also interacting with NS3. Earlier biochemical studies[148] also noted that subdomains 2 and the C-terminal subdomain of NS3 helicase are essential for interaction with NS4B.

1.3.2 Molecular Docking Simulation Shows That the N-Terminus Region of NS4B Interacts with RecA-2 and the CTD Subdomains of NS3

To validate our bacterial two-hybrid assay results and understand the interactions between NS3 helicase and NS4B, a molecular docking simulation was performed with a homology model built for NS3 of DENV serotype-1 and predicted the structure model of NS4B.

The crystal structures of NS3 are published for DENV serotypes 2 and 4. However, no crystal structure is available for the NS3 of DENV serotype-1. So a homology model was generated for serotype-1 NS3 using the DENV serotype-2 full-length NS3 (PDB Id. 2WHX) structure model as a template. The generated homology model was validated using the provided scores, and the model from the swissmodel server, the QMEANDisCo score for the predicted model is 0.81 + 0.6, and the model has 0.84% Ramachandran outliers.

There is no crystal structure model available for NS4B, and attempts at crystalization of NS4B protein resulted in bad-quality crystals which are not diffracted. Secondary structure prediction on the amino acid sequence predicts seven long helices that span through the majority of the protein sequence, except for the N-terminal 37 residues (Figure 1-3). Transmembrane helix predictions using different prediction software are inconsistent. TMHMM server did not predict any transmembrane helices, whereas PSIPRED MEM-SAT-SVM predicted six transmembrane helices. However, three TM helices are consistently predicted in most servers: the regions 100–150, 175–200, and

218–241 by Phobius and PSI-PRED MEMSAT-SVM prediction algorithms (Figure 1-3).

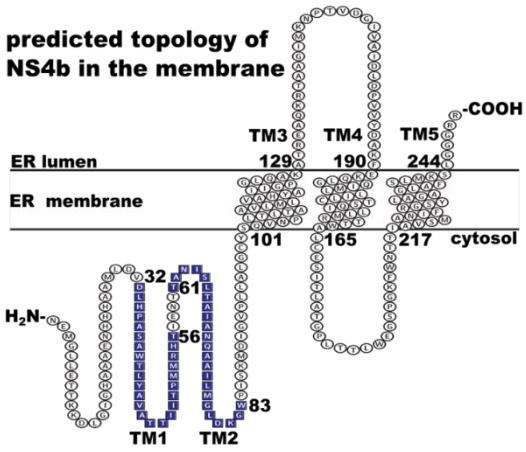


Figure 1-3 NS4B predicted transmembrane topology map (prepared using phobius)

The sequence of NS4B is unique, so there is no possibility of generating a homology model. RoseTTAfold and AlphaFold 2.0 use a machine-learning approach and are shown to accurately predict novel protein structures even when a homologous protein structure template is not available. The structure predicted by both AlphaFold and RoseTTAfold for DENV-1 NS4B, and these two models have identical folds when aligned RMSD deviation is less than 1Å (Alphafold is shown in Figure 1-4), so we proceeded with

AlphaFold model. In the AlphaFold model, for most parts of the sequence, except for the first 32 amino acid residues, the predicted local distance difference test (pLDDT) score is >90, which implies that the level of confidence in accurate prediction is high for that region (high likelihood of the predicted structure matching with the experimentally derived structure with low Cα-RMSD value). The N-terminus 32 residues region has a pLDDT measure of <70. The N-terminus 36 residues are predicted to be a disordered region, except for a small helix (LLETTKKDL) in the beginning (Figure 1-4).

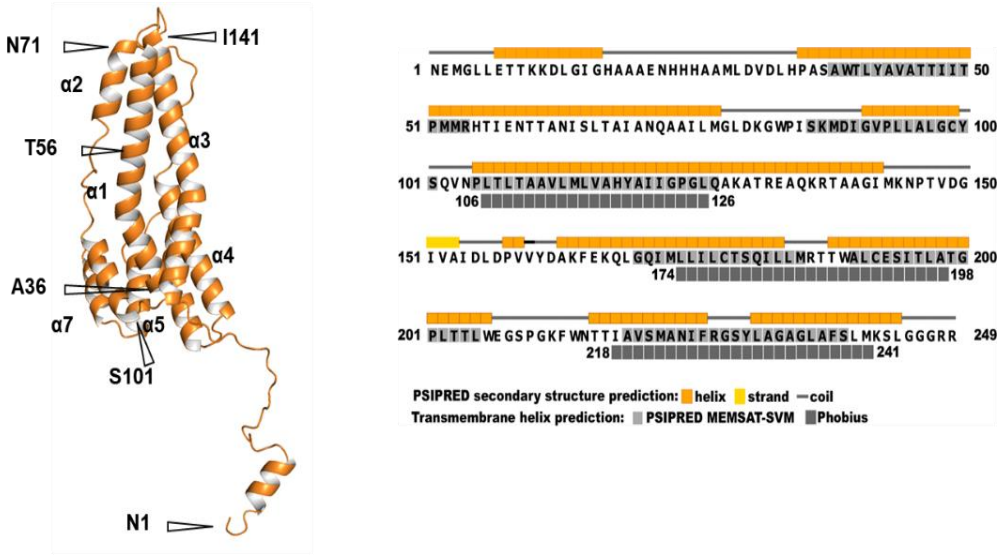


Figure 1-4 Ab-initio AlphaFold 2.0 structure model of NS4B, with schematic of sequence secondary structure and TM predictions

The predicted 3D structure of NS4B is in line with the secondary structure predictions on the sequence. A model for DENV serotype-1 NS3 using AlphaFold2.0 is also generated, explicitly not using any PDB template. Most of the sequence in the predicted model has a pLDDT score of >90, except for the N-terminal residues of the protease domain and the long loop connecting the protease to the helicase. The structural alignment of the model predicted by homology modeling and the model predicted by AlphaFold 2.0 model (only NS3 helicase domain) showed Cα-RMSD value of 0.886 Å, indicating that the AlphaFold

2.0 prediction is almost as efficient as model prediction by homology modeling. This increased our confidence in the model that we built for NS4B using AlphaFold2.0.

Later, a molecular docking simulation was performed with the generated structure models of NS3 and NS4B. The best-docked pose predicted by the Haddock server for the NS3–NS4B complex, as assessed by the lowest intermolecular energy score of -151.8 + 16.6 with a cluster size of 10 (Figure 1-5). As expected, only the N-terminal disordered region of NS4B exclusively interacts with NS3. The interaction interface spans recA-2 and the C-terminal subdomains of NS3; earlier studies also showed that NS4B interacts with recA2 and CTD of NS3[148]. The N-terminus 32 residues disordered region of NS4B runs lateral to the RecA-2 domain, on the side opposite the RNA-binding cleft, and reaches the C-terminus subdomain of NS3. This explains BACTH results wherein the observed interaction of full-length and N-terminal 57 residues region with all NS3 variants tested. In the docked position, Lys10 of NS4B and Glu568 in the C-terminal subdomain of NS3 helicase are at a distance it may form a salt bridge in between, and possibly stabilizing the interaction (Figure 1-5, zoomed-in region of the interaction interface). Notably, residues 10 to 28 in the NS4B are positioned next to a loop that connects the recA2 and CTD in NS3. This loop region exists as a helix in the apo (before ssRNA and ATP binding) form of NS3. This region is proposed to be a hinge (Figure 1-5) for the CTD swiveling motion on the recA2 after substrate binding[107]. Interestingly, when only the N-terminal 51–83 residues of NS4B were modelled and docked onto the NS3 helicase, the peptide model was positioned between the recA domains[168].

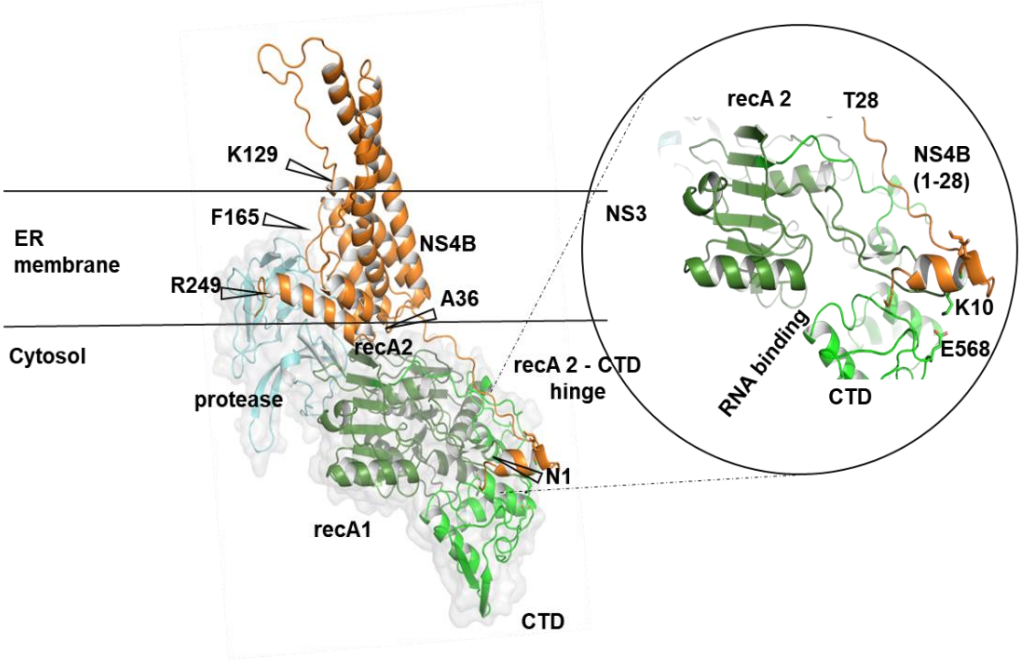


Figure 1-5 Molecular docking simulation : Haddock docked pose of NS3-NS4B (zoomed out image of interface of NS3-NS4B interaction and a putative salt bridge which may stabilize the interaction)

The structure model generated through AlphaFold is also marked with the possible boundaries of the membrane (Figure1-5), depicting the possible orientation of the NS4B in the ER membrane. If this were to be the orientation of NS4B TM helices in the ER membrane, then the N-terminal 36 residues disordered region of NS4B would extend as a flexible tail away from the membrane into the cytosol side of the RC and interacts with NS3. This model is in accordance with the mechanistic model of NS4B–NS3 interaction that others[141,179] have proposed—only the N-terminus flexible tail of NS4B would extend away from the membrane surface and interacts with NS3.

1.3.3 Recombinant NS4B and NS3 Proteins Interact Independent of NS4B Membrane Insertion

To further validate our BACTH and molecular docking analysis, DENV serotype-1 NS3 and NS4B proteins were expressed and purified, and thorough biochemical characterization of the interaction between the proteins was conducted.

The NS3 helicase domain was cloned (as per DENV serotype-1 sequence numbering residues 178–619 of NS3) into a bacterial expression vector in fusion with a C-terminal 6-histidine purification tag (Figure 1-6A). The bacterial two-hybrid assay and molecular docking results indicated that only the helicase domain interacts with NS4B. Moreover, earlier studies[149] showed that the bacterially expressed helicase domain folds into a functional protein independent of the protease domain of NS3. The purified protein elutes in gel-filtration chromatography as a single sharp peak at an elution volume that corresponds to the monomer mass of the protein (Figure 1-6B). Secondary structure analysis by far-UV circular dichroism spectroscopy shows a spectral signature (Figure 16C) typical of a predominantly helical protein, consistent with the earlier reported [107] secondary structure content for the protein.

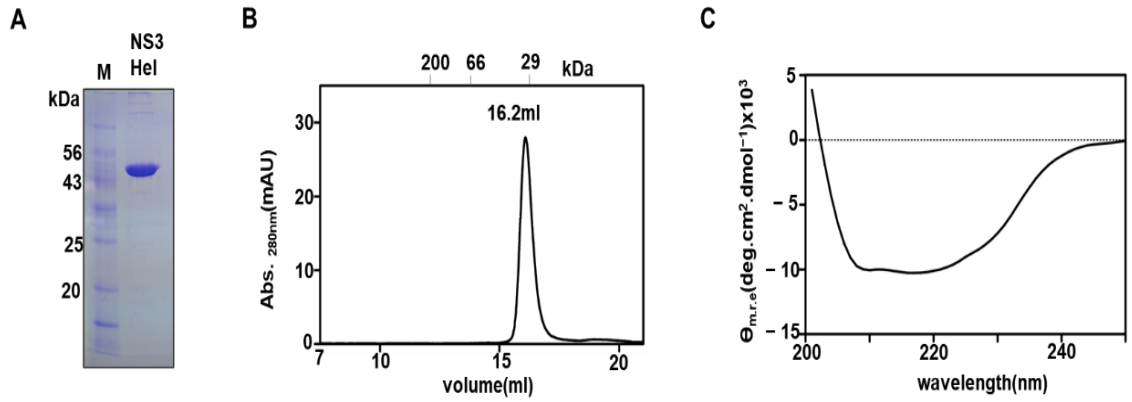


Figure 1-6 Recombinant NS3 helicase expression and purification (A) SDS-PAGE analysis of NS3 helicase (B)Size exclusion chromatogram of NS3 helicase (C) Far-UV CD spectra of NS3 helicase

The NS4B protein was also expressed and purified by adopting a protocol that was published earlier. Since NS4B is predicted to be a transmembrane protein, using LDAO detergent extraction of the NS4B protein was done from the insoluble fraction of bacterial cell pellets and purified using the 6-His Ni-NTA affinity purification protocol. To rule out the possibility of artifacts in the interaction studies due to the detergent, the protein was

exchanged into a buffer containing LDAO at a critical micellar concentration as the final purification step. SDS-PAGE analysis on the purified protein showed that the protein was more than 95% pure (Figure 1-7 A).

Size-exclusion chromatography on the purified NS4B revealed that the protein forms large oligomers, which are in equilibrium with a dimer and monomer population (Figure 1-7B). Two elution peaks were observed in the resolution range of the gel-filtration column and another broadly spread elution pattern, starting from the void volume and extending into the column resolution range. While the protein elution peak at 15.5 mL on a ~24 mL Superdex-200 column can be ascribed to the monomer (estimated molecular weight of 27 kDa; A 29 kDa marker protein, carbonic anhydrase, eluted at 16.5 mL), the peak at 14.2 mL corresponds to a dimer molecular mass of NS4B (Figure 1-7B). In the broad peak, the precise oligomer size of the protein, however, could not be determined. All three fractions contained only NS4B and not any contaminant proteins (Figure 1-7B, inset on the right side). The oligomers could not resolve even in a wider-resolution-range SEC column, the superose-6 column. Furthermore, the equilibrium shifted to large oligomers very rapidly, as the protein formed large oligomers within a few hours of storage at 4° C (Figure 1-7B, inset in the left). Interestingly, the oligomers of NS4B are entirely soluble, and there is no significant precipitation of the protein either on the gel-filtration column or in storage. Others made similar observations with purified recombinant NS4B[127] and the protein expressed in DENV-infected cells[149,167]. Far-UV CD spectroscopy (Figure 1-7C) on the protein reveals that the protein is predominantly helical (negative peaks at 208 nm and 222 nm), as expected from the secondary structure and tertiary structure predictions.

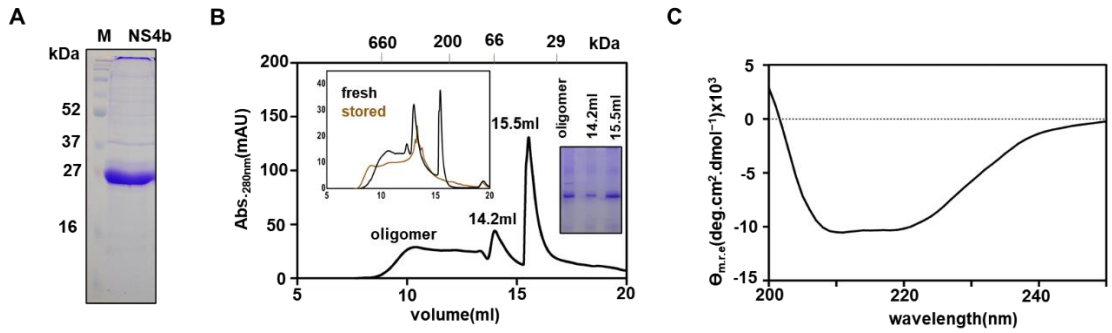


Figure 1-7 Recombinant NS4B expression and purification (A) SDS-PAGE analysis of purified recombinant NS4B protein (B) Size exclusion chromatogram of NS4B (C) Far-UV CD Spectra of NS4B protein

To characterize the interaction between NS3 and NS4B, the trials to isolate the monomeric fraction of the NS4B protein were made, as non-specific interaction between the protein molecules in the oligomers may mask the region that interacts with NS3. Despite several attempts and testing different buffer conditions (different non-denaturing detergent concentrations), isolation of the monomeric protein was not fruitful. Earlier studies[161] showed that the NS4B protein expressed in mammalian cells after a recombinant dengue virus infection could be co-immunoprecipitated along with NS3. It is may possible that NS4B oligomerization does not affect its interaction with NS3. An alternate possibility is that NS4B–NS3 interaction may induce conformational protein changes to populate more monomeric NS4B. An interaction kinetics assay was performed using Bio layer interferometry to test these possibilities and the interaction between the purified NS3Hel and NS4B proteins. The estimated KD value of the binding equilibrium is 0.508 μM (Figure 1-8A). These observations suggest NS4B oligomerization does not affect its interaction with NS3.

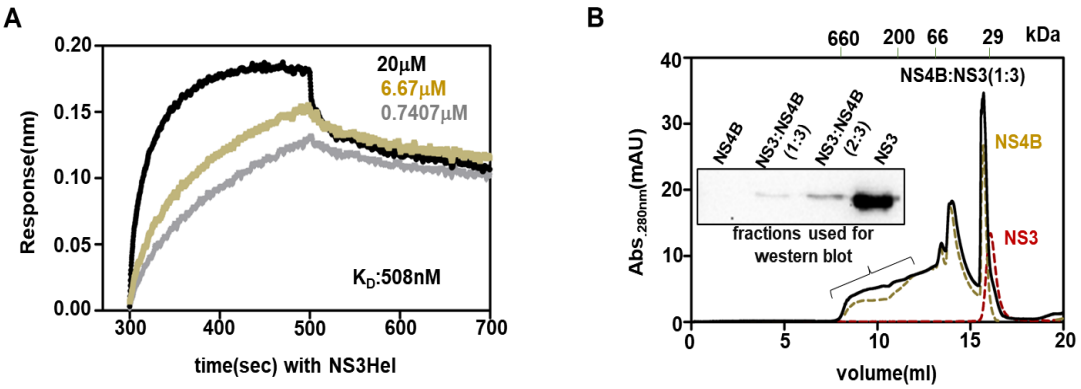


Figure 1-8 NS3-NS4B interaction with BLI and SEC (A) Bio Layer interferometry of NS3-NS4B interaction (B) Size exclusion chromatogram NS3-NS4B protein mixture

Mixtures of NS3Hel and NS4B were incubated and loaded onto size-exclusion chromatography at different stoichiometries to check the complex formation (Figure 18B). Fractions collected from each run were analyzed by Western blotting using an NS3specific polyclonal antibody. Surprisingly, there is no interaction found between NS3Hel with monomeric NS4B, as there is neither a significant reduction in the NS4B monomeric peak (green dotted lines in Figure 1-8B) nor the appearance of a new elution peak that would correspond to NS3Hel–NS4B (1:1 molecular ratio) molecular mass. However, the elution peak corresponding to the NS4B oligomers showed a small but significant increase in absorbance (280 nm) intensity (Figure 1-8B). When the fractions corresponding to this peak were analyzed by Western immunoblotting, a small amount of NS3Hel could be detected in those fractions, confirming that NS3Hel co-eluted along with NS4B oligomers. In gel-filtration runs with NS3Hel alone, the protein did not elute in those fractions.

To test whether the oligomerization of NS4B or the membrane insertion of the transmembrane helices of NS4B is a prerequisite for NS3 interaction. It is possible that, upon membrane insertion, NS4B would present the N-terminal 36 residues in the right conformation to interact with NS3, and oligomerization may provide a similar

environment. A co-flotation assay was performed with NS4B-lipid LUV mixtures, and NS4B did not show any association with lipid LUVs (Figure 1-9); NS4B is seen only in the bottom fractions. This observation negates our hypothesis that the membrane insertion of NS4B is a requirement for its N-terminus interaction with NS3.

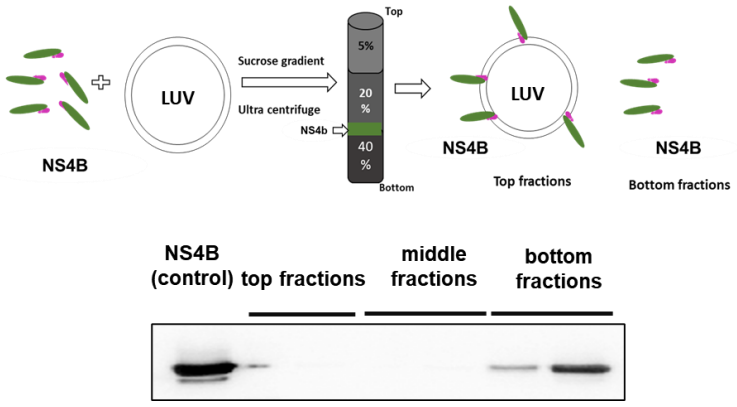


Figure 1-9 Lipid co-floatation assay of NS4B with LUVs, top- the schematic of lipid co-floatation assay, bottom is the western blot of lipid co-floatation assay top and bottom fractions with Anti His tag antibody.

These results show that the membrane insertion of NS4B is not necessary for its interaction with NS3 in vitro. A meaningful interpretation of these results is that the N-terminus disordered region of NS4B is enough to interact with NS3, as indicated by our bacterial two-hybrid assays and molecular docking studies. Furthermore, the oligomerization of NS4B does not hinder or effect the protein N-terminal region interaction with NS3.

1.3.4 Interaction Between NS4B and NS3 Leads to Conformational Changes in the Helicase Subdomains

To explain the mechanism of NS3 helicase activity modulation by NS4B, we tested if the interaction between NS3 and NS4B leads to conformational changes in the proteins. FarUV CD spectra of NS3Hel alone or an NS3Hel–NS4B protein mixture at different stoichiometries were recorded to test the secondary structure changes upon interaction. The blank-corrected and normalized spectra overlay (Figure 1-10B) is analyzed to

interpret the interaction described by Greenfield NJ, 2015[177] (see the Methods section for details). The far-UV CD spectra of NS3Hel and NS4B nearly overlapped. In the normalized spectra at higher ratios of NS3:NS4B, there is no significant change in the spectra of NS3, implying that there are no significant secondary structure changes in NS3 helicase upon interaction with NS4B if there are any secondary structure changes, a significant change in the spectra for the NS3Hel: NS4B mixtures is expected upon interaction.

Suppose the interaction results in large domain movements or changes in the tertiary structure; changes in the intrinsic tryptophan fluorescence can report such changes in the tertiary structure of the protein. Intrinsic tryptophan fluorescence of NS3Hel in the presence of different molar excess ratios of NS4B was recorded and compared it to NS3Hel alone spectrum, as can be seen in Figure 1-10C, the fluorescence intensity at the emission maximum (λmax emission: 345 nm) of NS3Hel in the presence of NS4B is high compared to NS3 alone. This signifies a change in the tryptophan environment in NS3 after the interaction. Nine tryptophan residues exist in the DENV serotype-1 NS3 helicase domain, of which five are in the C-terminal subdomain (Figure 1-10A). In the docked pose of NS3–NS4B (from our molecular docking studies), the N-terminal disordered region of NS4B arches around the RecA-2 and C-terminal sub-domains. Thus, we interpret the change in tryptophan fluorescence observed in our experiment as due to the N-terminal domain of NS4B interacting with the C-terminal subdomain of NS3 and possibly moving it closer to RecA domains. This would rigidify the environment surrounding the five tryptophans in the CTD, possibly resulting in the observed intrinsic tryptophan fluorescence changes.

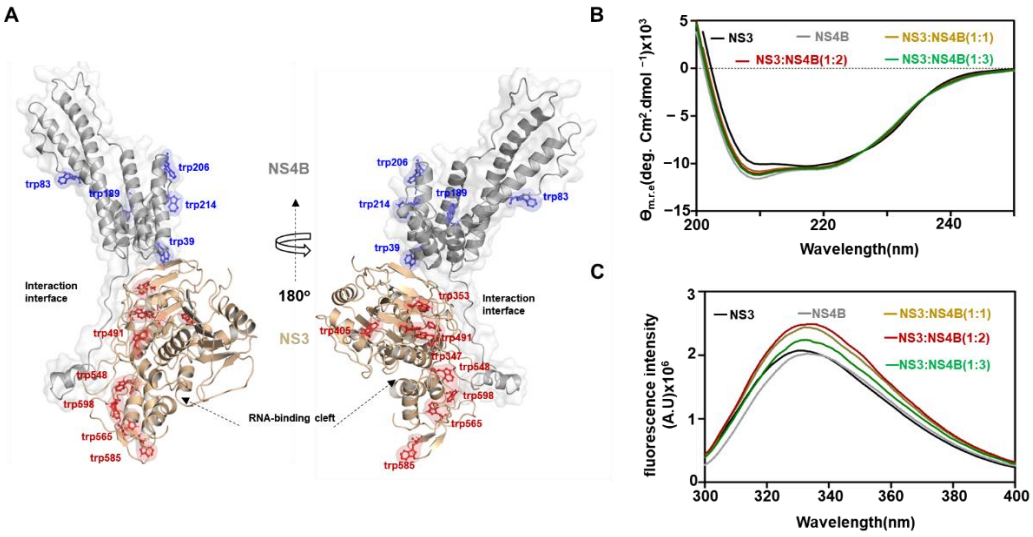


Figure 1-10 NS3-NS4B interaction biophysical characterization, (A) Schematic cartoon representation of NS3-NS4B Haddock docked pose with 'trp' residue marking in the interface (B) Normalized far UV-CD spectra of NS3-NS4B in different molar ratios (C)Normalized intrinsic tryptophan fluorescence spectra of NS3-NS4B in different molar ratios

Earlier molecular dynamics simulation studies on different flavivirus NS3 structures showed that, upon ATP-binding and hydrolysis, there are allosteric changes in the RNA binding cleft (between RecA domains and CTD)[105,180]. Based on this, it was explained that the allosteric changes possibly couple the RNA unwinding to ATP hydrolysis. The hinge connecting the CTD to the recA2 domain is flexible, and the CTD closing-in and moving-out motions into the RNA-binding cleft may contribute to the duplex-unwinding after each ATP hydrolysis cycle. BACTH and molecular docking studies showed that NS4B interacts with the CTD of the helicase.

Molecular dynamics simulations were performed on the NS4B-docked structure of NS3 to see if this interaction leads to any motion in the CTD, as it was interpreted from our intrinsic tryptophan fluorescence studies. An overlay of the end of the MD simulation trajectory structure with that of the initial structure (input of MD production run) is shown in Figure 1-11. As can be seen from the overlay of NS3 structures (NS4B is not shown in the Figure for clarity), the RecA and the protease domains did not show significant

motion. There was no significant movement of the center of mass of these domains during the entire trajectory. However, a center-of-mass analysis on the CTD showed that, compared to the initial structure (t = 0 ns in Figure 1-11A), the t = 20 ns structure moved by 3.3 Å, away from the RecA domains. An RMSD and RMSF analysis of the trajectory also showed a major deviation in residues corresponding to the CTD within the time-scale of the simulation. These results also suggest that CTD domain motions, probably on the hinge connecting the RecA-2 domain to the CTD domain, are possible away from and towards the RecA domains. As proposed earlier, the CTD domain motions may contribute to the RNA duplex unwinding activity of the helicase[170,180]. Furthermore, as revealed by our MD simulation results, the interaction of NS4B with the helicase CTD brings about similar motion in the CTD. However, we note that an all-atom or coarse-grained MD simulation for longer time-scales (microsecond scale) is required for reaching convergence of the simulation for a protein of the size of DENV NS3 (nearly 620 amino acids long). Future work in the lab aims to perform the simulations on a high-performance computer cluster and elaborate on the preliminary findings of this MD simulation study.

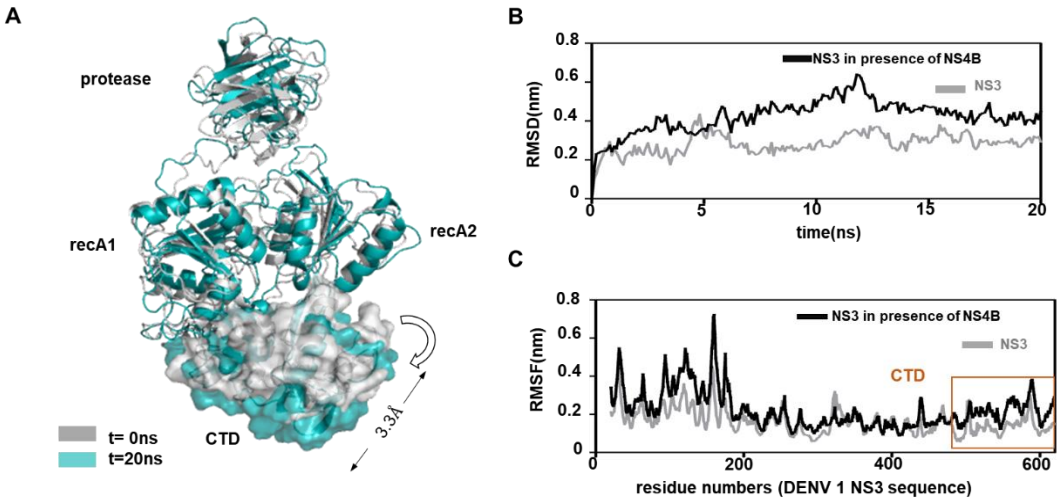


Figure 1-11 Molecular Dynamics simulation of NS3-NS4B (A) Trajectory snap shots at 0 ns(grey) and 20 ns(teal) with the surface representation of NS3 CTD subdomain (D) RMSD of NS3 with and without NS4B (C) RMSF of NS3 with and without NS4B (inset is showing the region of NS3 CTD subdomain)

To provide more direct evidence for the CTD domain motion into the RNA-binding cleft, upon NS3–NS4B interaction, crystallization of NS3 with full-length NS4B and NS4BN57 with NS3Hel was attempted. However, despite several crystallization attempts, co-crystallizing NS4B full-length orNS4BN57 along with NS3Hel. The crystals are obtained only with NS3Hel (Figure 1-12).

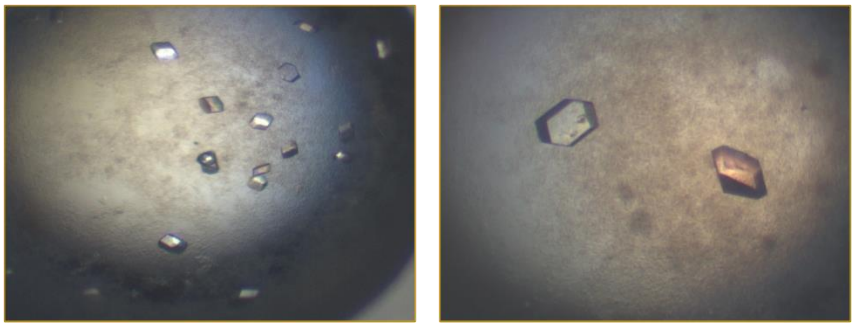


Figure 1-12 NS3 helicase protein crystallization, crystals of NS3 helicase

Attempts to express a smaller fragment of NS4B N-terminus (residues 1–36) or an N-terminus disordered region along with scaffolding proteins are ongoing in the lab. For a single molecule FRET experiment to see the domain motions in the presence of NS4B was planned and attempts to introduce a FRET pair of fluorophores between the CTD and recA2 through the introduction of cysteine or the introduction of an unnatural amino acid that is a photoactivable cross-linker in the CTD lead to protein misfolding (very low solubility). Attempts to express CTD alone were unsuccessful. All of these observations suggest that CTD plays a vital role in the structure and function of the helicase, and mutations in the subdomain may not be tolerated.

1.3.5 NS3 Helicase Activity on the 3′-SL of the 3′-UTR Is Enhanced Upon NS4B Interaction

Similar to many other SF2 helicases, flavivirus helicases show very poor in vitro helicase activity until an interacting protein enhances their activity. For dengue virus NS3 helicase, interaction with NS4B enhances its in vitro helicase activity[149]. Does the interaction between NS4B and NS3 helicase CTD cause helicase activity enhancement? A helicase assay was optimized for NS3 in the presence of NS4B to test the enhancement, using either an RNA or a DNA oligo corresponding to the 3′-SL regions of the DENV serotype1 sequence. Secondary structure predictions on the sequences show that both RNA and the DNA oligos can form a stable stem-loop structure (−29.00 and −15.83 kCal/mole of folding energy, respectively, as per RNA structure software prediction) that mimics the natural substrate of the helicase (Figure 1-13).

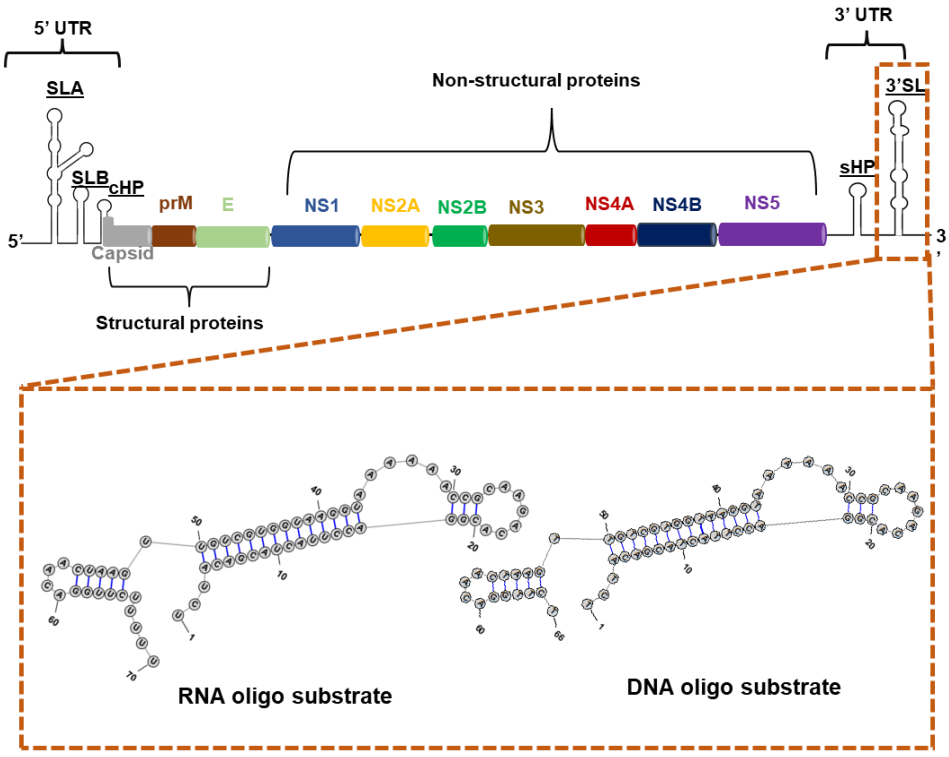


Figure 1-13 3'-SL representing RNA and DNA oligos substrates used for Gel based helicase assay

NS4B was purified without using LDAO detergent in the protocol for helicase assays. NS4B protein was purified without the detergent for the following reasons: (i) the LDAO detergent present in the NS4B preparation (purified following LDAO detergent extraction protocol) may interfere with NS3 helicase activity, (ii) And the results from this study showed that NS4B membrane insertion is not necessary for interaction with NS3, and (iii) the dynamic oligomerization of NS4B is observed even in the presence of detergent micelles.

NS4B purified without using detergent extraction showed far-UV CD spectra that almost overlapped with that of the protein purified in the presence of LDAO (Figure 1-14A) and formed large soluble oligomers (Figure 1-14B). However, we did not see a mono-mer or dimer population. This implies that there are no significant differences in the overall folding of NS4B with or without the membrane environment.

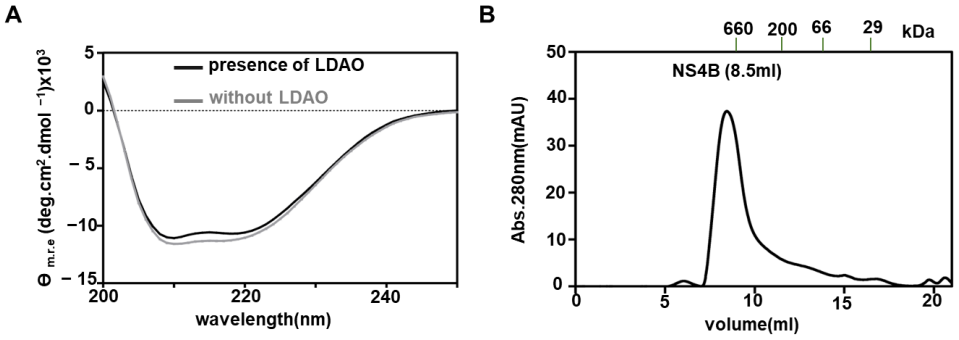


Figure 0-14 Secondary structure analysis of NS4B purified by Gd.Hcl (A) Far UV-CD spectra overlay of NS4B purified with and without detergent (B) Size Exclusion Chromatogram of NS4Bprotein purified without detergent

The N-terminal 57 residues region of NS4B (NS4BN57) was also expressed and purified in soluble form. Surprisingly, soluble NS4BN57 also oligomerizes, eluted at a volume corresponding to ~340 kDa (monomer mass is 7.5 kDa) on a Superdex-200 column. Gel-

filtration chromatography on a mixture of NS4BN57 and NS3 (1:1 molar ratio) showed that the proteins formed a stable complex. NS3 coeluted along with NS4BN57 oligomer (Figure 1-15 and inset in the right panel). This observation further validates our BACTH analysis and docking prediction that only N-terminal 57 residues of NS4B is enough to interact with NS3.

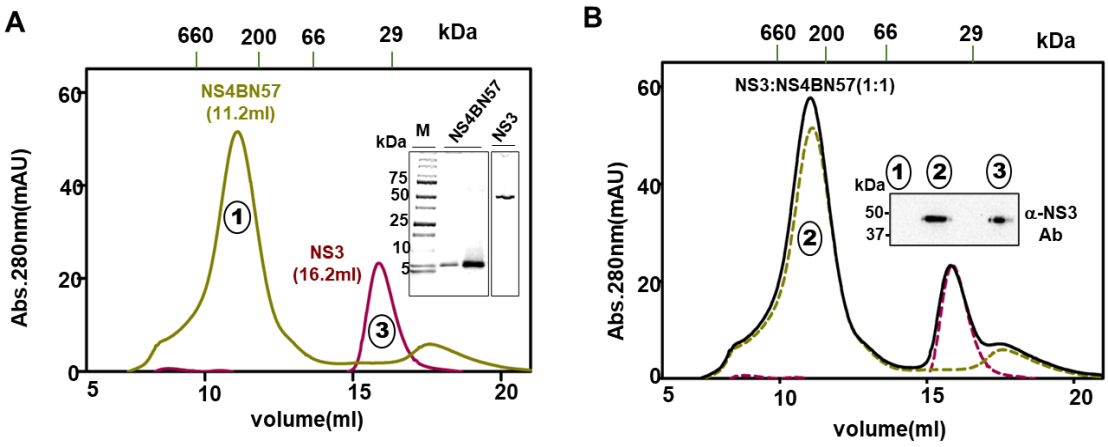


Figure 1-15 SEC analysis NS4BN57 and its interaction with NS3(A) Size Exclusion Chromatogram of NS4BN57 and NS3 independent control runs with inset image of peak fractions SDS-PAGE (B) Size Exclusion Chromatogram of NS4BN57:NS3 in 1:1 molar ratio with inset image of westernblot of peak fractions probed with polyclonal NS3 antibody

Either the full-length NS4B or the NS4BN57 proteins were used to test their effect on the helicase activity of NS3Hel.

Then, using the 3′-SL forming RNA and DNA oligos as substrates and purified NS3 and NS4B proteins, the NS3 helicase enzyme assay was optimized, adapting a gel mobility shift assay described earlier[95]. The rationale of the assay is that the stem-loop structure is more retarded in the electrophoretic gel compared to the resolved single-stranded oligo(Figure 1-16A). DENV NS3 helicase can use duplex DNA or duplex RNA as a substrate in vitro. In a control experiment, a sample of the heat-denatured DNA oligo (without allowing annealing) was loaded next to a sample of the annealed 3′-SL oligo in a gel-mobility shift assay. The annealed oligo is retarded more (ran above the denatured

oligo) than the denatured (completely unwound) oligo (Figure 1-16B), confirming that the 3′-SL oligos that we used formed stem-loop structures as predicted. Adding NS3Hel protein alone to the 3′-SL oligo did not show a band corresponding to unwound (Figure 1-16C, Lane 2), signifying that are no unknown components in the NS3 helicase protein preparation that may cause the unwinding of the stem-loop. Figure 1-16A shows a representative gel image from the helicase stem-loop unwinding experiment using the DNA oligo as substrate. In the presence of NS3 and ATP, a significantly lower band corresponding to the unwound 3′-SL DNA oligo is seen (marked ssDNA in Figure 116C). A band corresponding to the resolved DNA (ssDNA) is seen when NS3 is present in the assay mixture. No ssDNA band was seen when only NS4B was added to the assay or when ATP was omitted. This confirmed that the designed DNA oligo formed a stable stem-loop structure, as predicted, and the DNA duplex unwinding activity observed is precise because of the NS3Hel helicase activity. Two hours after the assay start time (time of addition of ATP to the reaction mixture), most of the stem-loop band intensity is reduced, and the intensity of the band corresponding to the resolved DNA increased. For a comparison of the helicase activity of NS3Hel in the presence of NS4B or NS4BN57, the one-hour time point was chosen as the endpoint of the assay, as this allowed us to assess the in-crease in helicase activity better.

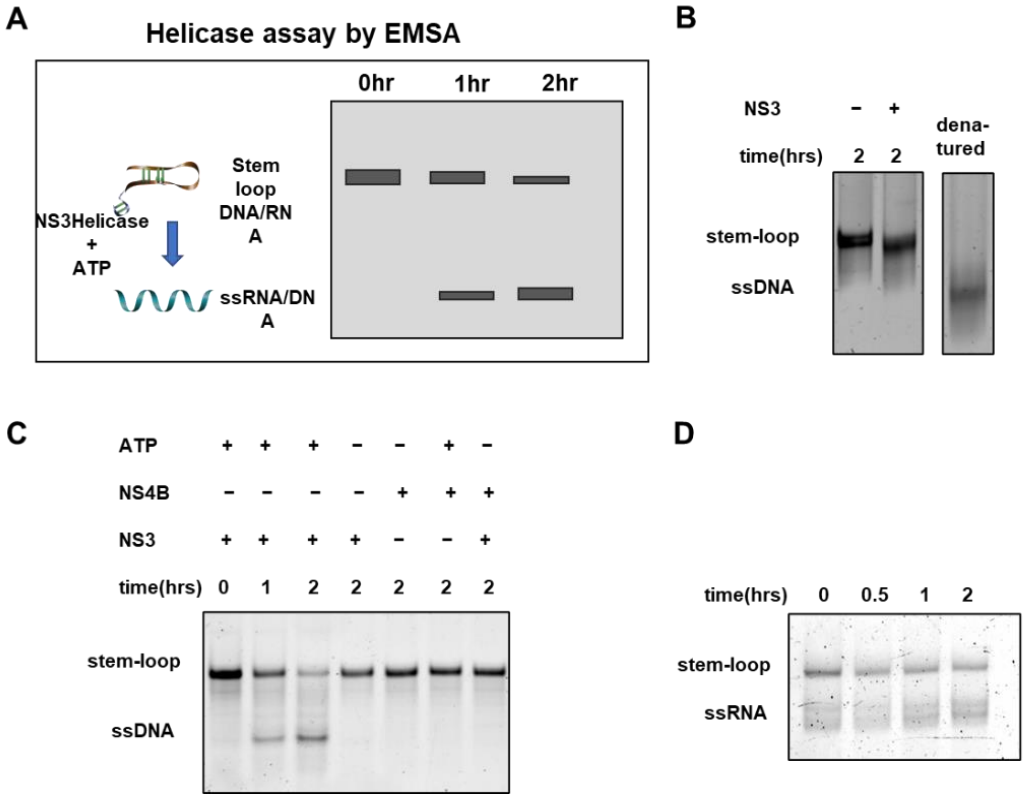


Figure 1-16 Helicase assay of NS3 (A) Schematic of NS3 helicase assay (B) EMSA of stem-loop DNA and denatured stem-loop DNA (C) EMSA on the DNA stem-loop substrate. As indicated, a band corresponding to the stem-loop resolved DNA (marked ssDNA) is seen only when NS3Hel and ATP are included in the assay. In the presence of NS3Hel alone (no ATP) or NS4b alone, the DNA stem-loop (D) EMSA of RNA oligo-based NS3 helicase assay

NS3 helicase assay was also performed using 3′-SL RNA oligo as substrate with FAM label at 5’ side—the natural substrate of the helicase. A representative image of the gel from the gel mobility shift assay is shown in Figure1-16D. Unlike the 3′-SL DNA oligo, the RNA oligo showed a few lower bands in addition to the band corresponding to the stem-loop. However, when the oligo was incubated with NS3Hel and ATP in the assay buffer, the band intensity corresponding to the unwound oligo (marked ssRNA in Figure 1-16D) increased, with a concomitant decrease in the intensity of the stem-loop band. This confirms that the NS3Hel can use the 3′-SL RNA oligo as a substrate and can unwind it.

The first in vitro biochemical evidence is that DENV NS3 helicase uses 3′-SL as a substrate. However, we noticed that when NS4B proteins were added to the assay mixture along with the RNA oligo, a major proportion of the oligo precipitated and did not enter the gel. This probably happened because the hydrophobic dye (FAM) that we conjugated to the oligo for detection after the electrophoretic run binds non-specifically to the hydrophobic patches on NS4B. As a result, we could not use the RNA oligo in the helicase assays with NS3–NS4B or NS3–NS4BN57 mixtures.

A fluorophore and a quencher attached single-stranded RNA oligo used in a molecular beacon assay to test DENV NS3 helicase assay adapting a protocol described earlier[106]. As schematically depicted in Figure 1-17, the fluorophore–quencher pair attached to an RNA oligo is annealed along with a complimentary un-labeled oligo. If NS3Hel unwinds the duplex RNA, the labeled oligo can form a stem-loop structure, thus bringing the fluorophore and the quencher into FRET distance, leading to quenching of fluorescence. The extent of quenching is used as a measure of NS3 helicase activity (one representative CY5 fluorescence intensity vs. time of the assay is shown in Figure 1-17, next to the schematic). The sequence of the labelled-RNA oligo is not that of the 3′-SL region, though. This assay was used to measure the helicase activity on the RNA duplex in the presence of NS4B and NS4BN57 in place of the gel-mobility shift assay. In the presence of NS4B or NS4BN57, the helicase activity on the RNA duplex was enhanced (table in Figure 117). However, the fluorescence signal was very noisy when NS4B was included in the assay, resulting in large standard errors in the experiment. Since the 3′-SL RNA and DNA oligo as the substrate of NS3Hel did not show any difference, the 3′-SL DNA oligo as a substrate for further studies to quantify the helicase rate enhancement in the presence of NS4B or NS4BN57.

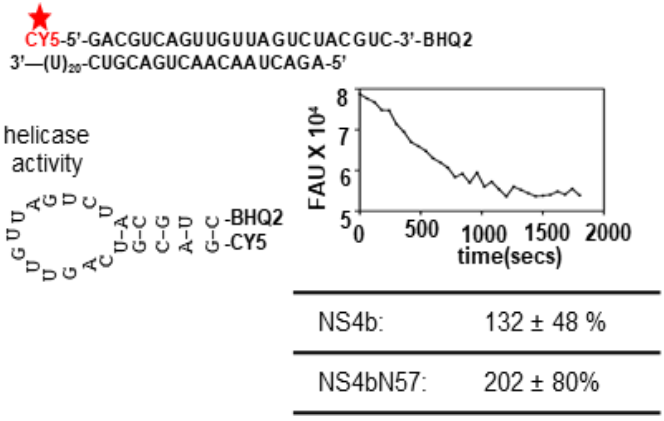


Figure 1-17 Molecular beacon-based helicase assay of NS3 in the presence of NS4B and NS4BN57 percent increase in helicase activity.

To assess the effect of NS4B on NS3 helicase activity, performed the helicase assay in the presence of different molar ratios in excess of NS4B to NS3Hel using 3′-SL DNA oligo as substrate. A representative gel image from the helicase assays is shown in Figure 1-18A. The ratio of background-corrected pixel intensities of the un-wound oligo band (ssDNA) to the sum of pixel intensities from the stem-loop band and unwound (ssDNA) band is expressed as % helicase activity (refer to the Material and Methods section for details). The activity of NS3Hel protein at the assay end point (1 h post addition of ATP into reaction mixture) is taken as 100% activity. If any, the pixel intensity of the unwound (ssDNA) band at the zero-time point (stopping the reaction within five minutes of adding ATP to the assay mixture) is subtracted in each case to normalize the zero-time point activity to 0%. In the presence of a three-molar excess of NS4B, NS3Hel helicase activity increased by 53 (SD of ±16) % over and above the NS3Hel-alone activity (Figure 1-18B). Similarly, in the presence of NS4BN57, the activity increased by 22 (SD of ±14) % (Figure 1-18C). Notably, though NS4BN57 showed interaction with NS3Hel at 1:1 and 1:2 molar ratios, the interaction did not increase the helicase activity of NS3Hel to the same extent as what full-length NS4B could bring about. It is possible that, though the N-

terminus 57 residues region of NS4B is enough to interact with NS3 helicase, other regions of the NS4B may be important for helicase activity modulation in ways that cannot be explained in this study.

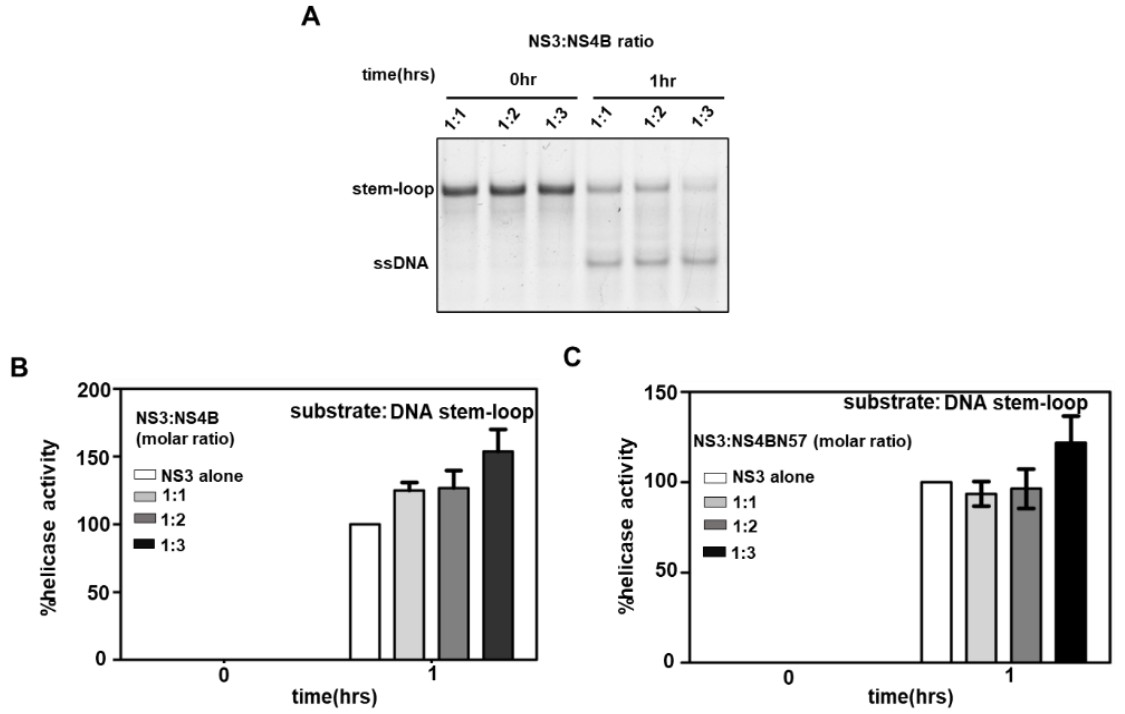
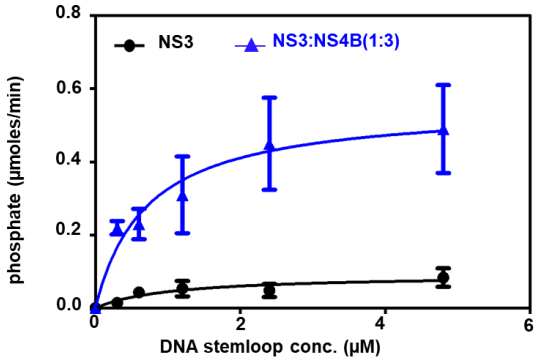


Figure 1-18 Helicase assay of NS3 in the presence of NS4B and NS4BN57(A) EMSA of NS3 helicase assay in the presence of NS4B (B) Histogram of a percent increase in helicase activity of NS3 in the presence of NS4B (C) Histogram of a percent increase in helicase activity of NS3 in the presence of NS4BN57.

ATPase activity assay was also optimized for the of the helicase and used it to determine the Michaelis–Menten kinetic constants of the reaction. The reaction initial velocities at different substrate concentrations fit an MM-kinetics model using a non-linear regression method as shown in Figure 1-19. In the presence of NS4B (at a 1:3 NS3 to NS4B molar ratio), the Vmax of the reaction increased to 0.560 + 0.051 μmoles/min from 0.092 ± 0.018 μmoles/min for the NS3Hel-alone reaction, implying a nearly 6-fold catalytic rate enhancement of the helicase upon NS4B interaction. Assuming that the ATPase activity

and RNA duplex unwinding activity are correlated, an increase in RNA unwinding activity of the helicase to the same extent (as ATPase activity) is expected.



	Km (µM)	V_{max} (µmoles/min)
NS3	1.038 ± 0.57	0.092 ± 0.018
NS3:NS4b	0.700 ± 0.20	0.560 ± 0.051

Figure 1-19 ATPase assay with NS3 helicase in presence of NS4B

Taken together, These results from gel-shift mobility assays using RNA and DNA oligos predicted to form a 3′-SL structure, molecular beacon assays with RNA oligos, and MMkinetic studies show that the 3′-SL in the 3′-UTR region of the DENV genome is a cognate substrate for NS3 helicase and the interaction of NS4B with NS3 helicase enhances the in vitro helicase activity. These results also validate our observations from BACTH assays and molecular docking studies that the N-terminus 57 residues region of NS4B is enough to interact with NS3 helicase.

1.4 Discussion

Flavivirus non-structural proteins interact with each other and with host proteins in a complex and dynamic manner[160,181]. Earlier studies established the criticality of NS3–NS4B interaction[148,161] in the DENV replication. Furthermore, a therapeutic strategy has been proposed to target the NS4B–NS3 interaction in DENV and other flaviviruses.

Recently, a small-molecule drug, JNJ-A07 (a hit from a large-scale cell-based anti-DENV 2 screens), has been shown to potently inhibit viral replication by targeting the NS4B–NS3 interaction[11]. However, the interaction between NS3 and NS4B, and its significance to the enzymatic functions of NS3, is not completely understood. With the objective of characterizing the interaction between NS3–NS4B and providing structural and mechanistic insight into the interaction, this study characterized the DENV 1 NS3–NS4B interface and explained a possible structural mechanism for the helicase activity enhancement of NS3 upon the interaction.

Earlier studies showed[149], using a yeast-two-hybrid assay and a pull-down assay on DENV-infected cell lysates, full-length NS4B interacts specifically with the NS3 helicase subdomains recA2 and CTD[148]. BACTH results and docking simulations are consistent with their results. However, in their assays, the N-terminal half (1–135) or the C-terminal half of NS4B (136–248) alone did not interact with NS3. Interestingly, NS3 also interacts with NS5 through CTD subdomain[147,182]. Both NS5 and NS4B may interact with NS3 subdomains two and three and form a tripartite complex during negative-strand synthesis in genome replication[181].

Docking studies in this study corroborate their findings. The binding strength of the interaction reported in their study (KD of 222 nM for DENV2 and 530 nM for DENV4 NS3Hel–NS4B) is comparable to what we observed in our BLI studies (KD of 508 nM). They also tested the interaction between a peptide with a sequence corresponding to a loop connecting the NS4B TM3 and TM4 with NS3, and their rationale for selecting this region is based on the predicted membrane topology of NS4B, wherein the TM3–TM4 loop is the only region that would face the cytosolic side of the RO, wherein NS3 is localized. Interestingly, the simulated KD values estimated for NS3–NS4B cytosolic loop

binding (1.6 μM) is very high compared to those observed with full-length NS4B and helicase and the same region involved in dimerization of NS4B[127,148]. This implies that there may be other regions of NS4B that can interact with NS3. Further, NMR studies with the cytosolic loop peptide and NS3 helicase domain, followed by genetic analysis, fine-mapped the interacting residues to Q134 and G140. Another independent study found that the Q134 residue of the TM3–TM4 cytosolic loop is the determinant of NS4B interaction with NS3[161]. Interestingly, in their studies, alanine substitutions at DENV2 NS4B residue positions 28 (L28A) and 87 (M87A) also showed the apparent loss of interaction with NS3, albeit to a much lesser extent than the cytosolic loop mutations (Q134A, G140A, and M142A).

The studies described above on mapping the NS4B–NS3 interaction interface looked for interaction between the NS4B TM2–TM3 cytosolic loop residues and NS3. Their rationale is, as per the predicted membrane topology of the NS4B [183], the 2K signal preceding NS4B would lead the protein into the membrane from the ER lumen side, and thus, the N-terminal 100 residues region is placed in the ER lumen. In that scenario, the cytoplasmic loop between TM3–TM4 is the only accessible region on the cytosolic side of the RO for NS3 interaction. This study has a differing view about the localization of N-terminal 100 residues of NS4B, that the N-terminal 100 residues are disordered and extend away from the membrane surface into the cytosol side of the RO (as depicted in Figure 1-3, schematic of the RO on ER) which would allow its interaction with NS3 and other interacting partners. The reasons for our proposal are as follows: (i) Miller et al. (2006)[167], based on TM helix predictions and biochemical studies, proposed a membrane topology for NS4B, which the earlier mentioned NS3–NS4B interaction studies were based on. However, the proposed topology model does not exclude the possibility of the

N-terminal region flipping to the cytosol side of the RC after proteolytic processing at the C-terminal end of the 2K signal peptide. This may happen similarly to what is proposed for the TM5 of NS4B—TM5 flips from the cytosolic side to the ER lumen side after the NS4B–NS5 boundary is cleaved[122]. Alternatively, after the cleavage at the 2K signal peptide end by host signalase, the N-terminus of NS4B may come out of the membrane to the cytosolic side, thus flipping the whole NS4B orientation in the membrane opposite to what Miller et al. proposed, (ii) Zou et al. (2014)[127], based on their fluorescence protease protection assay results with DENV 2K-NS4B(1-93)-EGFP, inferred that the N-terminus 100 residues might position on the ER lumen side, as well as on the cytosol side, of the RO; (iii) Zou et al. (2015) in their study could immunoprecipitate NS4B from DENV-infected cell lysates without the detergent extraction of the protein. Since the majority of the NS4B, except for the N-terminal 100 amino acids, insert into the membrane, the only explanation for this result is that the N-terminus is on the cytosolic side of the RC, where NS3 is also present; (iv) The N-terminal 95 residues of NS4B are required for the modulation of IFNα/β signaling[169] in DENV-infected cells, probably through its interaction with the STAT-1 protein, a cytosolic protein. Thus, in this study proposal, the N-terminus region of NS4B extends like a tail into the cytosol for interaction with various protein factors is not unfounded. Consistent with this idea, recently study[168] showed that the N-terminal 51–83 residues of NS4B are enough to interact with NS3 and enhance NS3 helicase activity. However, their study was done with peptides corresponding to different regions of the N-terminal region or the cytosolic loop connecting the TM3 and TM4 of NS4B (as Sumo-fusion proteins).

This study did not focus on any region of NS4B and NS3 to map the interaction between them, unlike the studies mentioned in the above paragraphs. The molecular docking

simulations were performed with full-length proteins (including the protease domain of NS3) without a prior notion about the interaction. Furthermore, in vitro studies with purified proteins (full-length NS4B or NS4N57 and NS3Hel) to characterize the interactions validated our docking studies. This study concludes that the N-terminal, likely disordered, region of NS4B is the major determinant for interaction with NS3. Based on these observations, a mechanistic model for the NS3–NS4B interaction was proposed (Figure 1-20). Though speculative at this juncture, it is reasonable to think that the N-terminal disordered region of NS4B flips from the ER lumen side to the cytosol side (by a mechanism that cannot be explained from our studies) after the C-terminal end of the 2K signal is proteolytically processed. Then, the N-terminal end extends as a tail down from the membrane surface, runs lateral to the recA-2 domain, and partially wraps around the CTD from the side opposite to the RNA binding cleft. As we noted in our docking analysis, a salt bridge between the K10 of NS4B and NS3 CTD E568 may stabilize the interaction to tightly hold the NS4B N-terminal region around the CTD. It would be interesting to see if mutating the residues, K10 from NS4B or E568 in NS3 would break the interaction between NS4B–NS3 when mutations introduced in this region the neighbouring residues to E568 in NS3 (for using them in crosslinking studies), the mutant proteins did not express well, possibly because of the misfolding of the protein. Thus, this new interaction interface between NS4B and NS3 that was proposed here, along with interaction with the loop regions between the TM helices proposed earlier, is likely to be the complete interaction interface between the two proteins. This complete map of the NS3–NS4B interaction interface should be considered in designing therapeutic strategies to block the interaction.

What does interaction with NS4B signify to the enzymatic functions of NS3? Since only the helicase subdomains are involved in interaction with NS4B, it is reasonable to assume that NS4B can influence the helicase activity of the protein. Consistent with this idea, earlier study[149] showed that GST-linked NS4B could enhance the duplex RNA unwinding activity of NS3 by nearly two-fold at a 1:2 NS3 to NS4B ratio; however, a mechanism that can explain the rate enhancement is not known. In this study, an RNA or a DNA stem-loop would be analogous to the natural substrate of NS3; the 3′- stem-loop in the 3′UTR was used. This stem-loop (3′-SL) is likely to be the first substrate for NS3 during negative-strand synthesis. Biochemical characterization of NS3 showed that the 3′-SL is indeed a cognate substrate for the NS3 helicase. It is important to note that earlier studies with NS3 helicase used either an RNA duplex or stem-loop forming sequences rather than 3′- SL from the 3′UTR of the viral genome.

Upon interaction, the conformational changes in NS3 and NS4B proteins were also studied. Though inconclusive, our intrinsic tryptophan fluorescence and MD-simulation studies suggested that NS4B–NS3 interaction may increase the dynamic motion of the CTD (subdomain 3) of NS3. Based on the crystal structures of the DENV NS3 helicase in different ligand-bound (an ATP-analogue, Mn^{+2}, intermediates of ATP hydrolysis, and a ssRNA fragment) states, Luo et al. (2008)[107] proposed a mechanism of NS3 nucleoside triphosphatase activity and the coupled dsRNA unwinding activity. These crystal structures reveal that ssRNA is held in the RNA-binding cleft by interactions with several residues in all three subdomains, although most interactions are with recA1 and recA2 subdomain residues (primarily with RNA back-bone phosphoryl oxygen). They also noted that, in the ssRNA-bound structure, subdomain 3 (CTD) rotates by 11 degrees, away from the RecA domains, to make the RNA-binding cleft wider. Based on these

observations, they proposed that subdomain motions after dsRNA binding in the RNA binding cleft (CTD opening and closing onto Rec-A subdomains) may provide the wrenching force to push the duplex through the 'separation pin'—a β-hairpin structure capping the RNA-binding cleft of NS3. A similar proposal—dynamic CTD (subdomain 3) motions during the unwinding action of helicase to close the RNA-binding cleft—was made for Zika virus NS3 helicase[170,180] and hepatitis C virus NS3 helicase[184]. Under this hypothesis, a molecular dynamics simulations study on the dengue virus NS3 helicase structure in complex with RNA and ATP hydrolysis intermediates predicts dynamic conformational changes in the protein, explaining the allosteric mechanism connecting the helicase activity to the ATP hydrolysis[185].

Interestingly, in the NS4B-docked NS3 structure, the disordered region of NS4B is closely apposed to the 'recA2-CTD hinge' of NS3. Furthermore, our MD simulation studies on the NS3–NS4B docked structure indicate CTD domain motions. Our study provides experimental evidence, though not direct evidence, supporting this proposed mechanism. Tertiary structure changes in the NS3–NS4B complex that we observed correlate well with the rate of enhancement of the helicase activity. At higher NS3 to NS4B molar ratios, wherein helicase activity is higher, the intrinsic tryptophan intensity significantly enhanced. This change can be interpreted as more rapid dynamic motions in the subdomains resulting in enhanced helicase activity.

Could the NS4B N-terminus interaction with NS3 RecA2 and CTD domains enhance the subdomain's motion to increase the helicase activity? The results of our molecular docking, far-UV CD and intrinsic tryptophan fluorescence studies, MD simulations, and helicase assays indicate such a mechanism. As described in the Results section (refer to Figures 1-5 and 1-10), it is reasonable to interpret the observed change in intrinsic

tryptophan fluorescence in the NS3–NS4B complex (but no far-UV CD changes) as primarily due to CTD moving closer to the RecA domains.

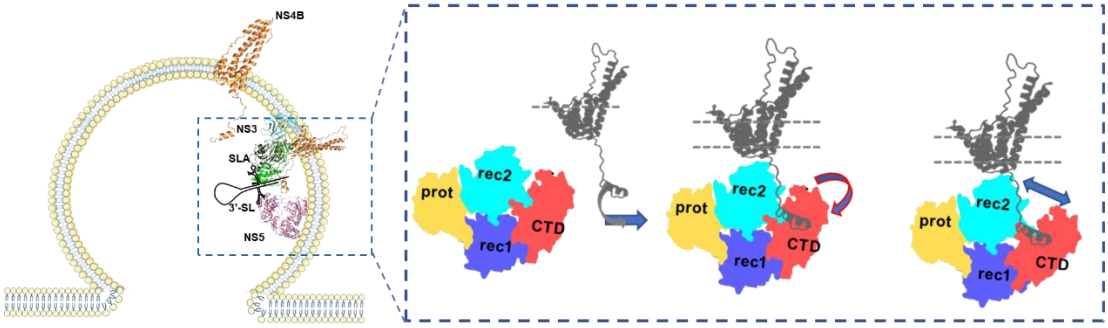


Figure 1-20 Plausible structural mechanism of NS4B induced NS3 helicase activity enhancement by inducing dynamic motion in CTD of NS3 by N-terminal disorder region of NS4B

Taken together, a plausible explanation is that NS4B interaction with NS3 helicase increases the RNA duplex unwinding activity by increasing the CTD motions (towards and away from the RecA domains for each cycle of unwinding/RNA translocation).

Chapter 2: Characterization of NS2B Co-factor Interaction with NS3 Protease

2.1 Introduction – DENV NS3 Protease and Polyprotein Processing

Dengue genome has a single ORF. In the infected cell, the viral genomic RNA is translated into a single long polypeptide which is proteolyticall processed by the viral and the host cell proteases to produce seven non-structural and three structural proteins. Mutations in these cleavage sites cause abrogation of polyprotein processing and has severe consequences for viral survival, majorly affecting its replication and virus particle synthesis[186–191]. DENV polyprotein is specifically cleaved at eleven sites by host signal peptidase, furin, and the viral-coded protease, NS3. Host signal peptidase cleaves specifically at the capsid-prM boundary, the prM-envelope protein boundary, the envelope protein – NS1 and the 2k-NS4b boundary[44]. Remaining six cleavage sites are recognized and cleaved specifically by the NS3. As described in the general introduction and chapter 1, the N-terminal 176 residues of DENV NS3 fold into a canonical chymotrypsin fold with two b-barrel domains. The NS3 protease (domain) has serine protease activity. The catalytic triad of Ser 135 - His 51 – Asp 75 (residue numbering as per DENV serotype 1 sequence) is formed at the boundary of the two domains.

Interestingly, (as described earlier) the NS3 protease requires a protein co-factor, NS2B (the non-structural protein that immediately precedes NS3 in DENV polyprotein sequence). The C-terminal (nearly 50 residues long) half of the NS2B protein folds around the NS3 protease domain to stabilize and functionalize the protease [151,192]. NS2B's role as a co-factor stabilizing the NS3 fold is well established, but only the region of 45-56 amino acids showed interaction in the 'open' conformation of NS3, which alone stabilizes the fold[87]. NS2B's structurally conserved beta-hairpin motif (73-86amino acids region) is positioned differently in both the conformations (Figure 0-9). In 'open' conformation away from the protease domain. In 'closed' conformation, the extended

region of NS2B co-factor (56-100aa region) forms two new interfaces (first interface formed by 71-75 residues region of NS2B, and 113-119 residues region of NS3 protease will interact with NS3 protease), and the conserved beta-hairpin motif is conformationally positioned near the catalytic pocket ('DD' in the loop region of this beta-hairpin motif interacts with '73KK74' residues of NS3 before the catalytic aspartate forms the second interface) this conformational rearrangement and the two interfaces formed will stabilize the closed conformation [193].

The extended region of NS2B cofactor from the 56-100aa region is essential for the NS2B3 protease activity, mutations in two different regions of this loop ranging from 5962 and 75-87 showed decreased proteolytic activity [15], at positions L75A, I77A, and I79A showed insufficient autoproteolysis, and a single mutant W62A completely abolished autoproteolytic activity on 2B/3 substrate[86,195]. Despite having significant sequence conservation within the flavivirus genus, the cofactor of each virus seems specific for its protease, although cross-activation within closely related viruses is also observed sometimes[151]. Although in the open conformation of NS2B3 protease, this extended loop region of NS2b3 is away from the core protease domain, the mutational studies showed this region is required for the activity of NS2B3 protease. These mutational studies and the conformational dynamics of NS2B cofactor region are clues about this region's criticality and the importance of closed conformation formation for the activity of NS2B3 protease.

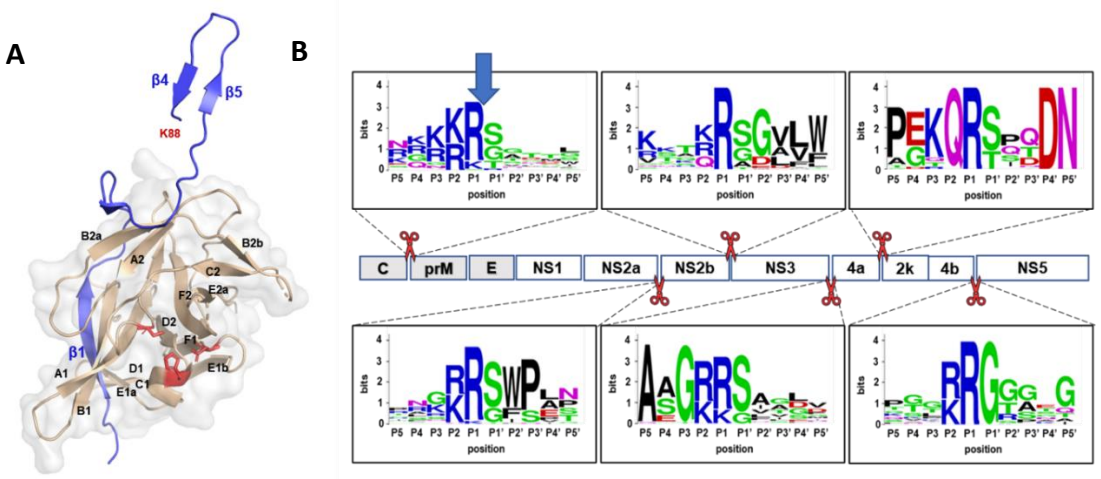


Figure 2-1, NS2B3 structure and its cleavage site (A) NS2B3 protease structure, (B) MSA generated Logo of conserved residues in the cleavage site of NS2B3 protease

Cleavage sites of NS2B3 protease in between viral proteins have a conserved pattern with slight heterogeneity throughout the flavivirus genus; the pattern of the cleavage site of NS2B3 protease has two basic residues followed by a small side chain having aliphatic residues like glycine or alanine or serine, or threonine (Figure 2-1B). The cleavage site always has R/K at P1 and P2 positions, although there is an exception for NS4A-2K and NS3-NS2B, where the P2 position is occupied by 'Q'; otherwise, in most cases, the P2 position is another 'basic' residue. However, in the case of NS4A -2K and NS2B-NS3 cleavage site, P2 is occupied by glutamine and P3 position occupied by another 'basic' residue, except in DENV3 where NS2B-NS3 cleavage site has P2, P3, and P4 positions occupied by glutamine, threonine, and Glutamine whereas P5 position is another 'basic' residue.

NS2B3 protease cleaves between this P1 'basic' residue and P1′ small sidechain having residue, where P1 'basic' residue side chain was held in the S1 pocket (Figure 2-2). The cleavage site recognized by NS2B3 protease has slight microheterogeneity in the same serotype. However, then also NS3 specifically recognizes this cleavage junction and

cleaves it at this boundary after basic residues raising questions about S2 pocket specificity for the P2 position. Moreover, how this second 'basic' residue slides between P2 to P5 in some of the cleavage sites recognized by NS2B3 indicates a broader pocket for the second 'basic' residue, which defines the cleavage site specificity.

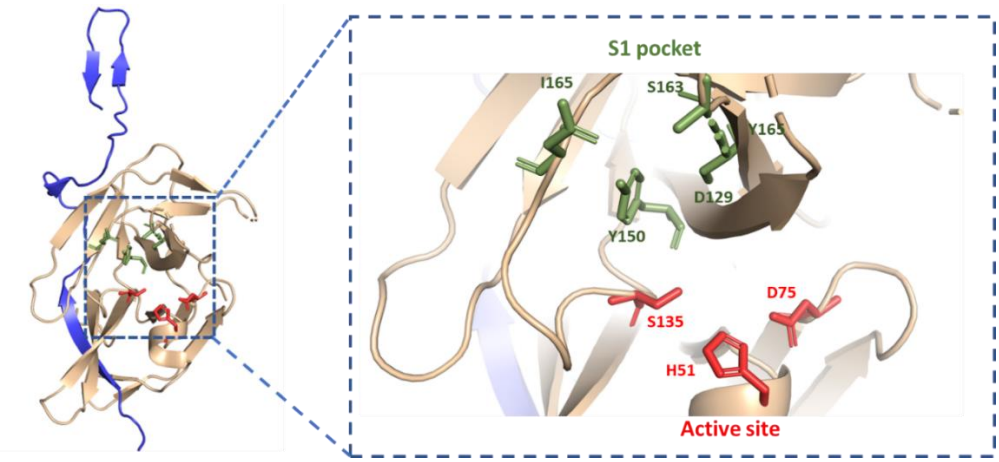


Figure 2-2 NS2B3 protease with the active site and S1 pocket residues

Comparison of other flavivirus proteases like HCV NS3 protease, which cleaves after cystine shares a similar fold, and its cleavage specificity is influenced by aspartate ate at P6 position[154].

Negatively charged residues in loop region beta-hairpin motif of NS2B in WNV made S2 pocket for P2 'basic' residue. However, in DENV, it was not clear, although an MD simulation study showed this negatively charged residue might interact with conserved 'basic' residue near catalytic aspartate stabilizes closed conformation for the activity of NS3[196,197]. Other naturally occurring proteases like trypsin and thrombin, NS2B3 also has an aspartate six residues before catalytic serine, which gives cleavage specificity after 'basic' residue in trypsin and thrombin which holds side chain on this 'basic' residue during peptide bond cleavage, but in NS3 mutation at this position has minimal effect[155].

Upon analysis of different flaviviruses like ZIKV, WNV, and different serotypes of DENV protease showed a dimer form of NS2B3 protease in 'open' conformation of crystal structure, and whenever it is bound to substrate, or an inhibitor showed a monomeric 'closed' conformation mostly[198]. This indicates that its conformational change during substrate-bound form may be a stimulus for its activation. NS2B is forming a complex with NS3, and its cofactor role of stabilizing the fold was well established. However, its role in protease activity is not a well-known phenomenon. Along with this, there are a few questions like how NS2B3 got its specificity to cleave at specific sites in between structural and nonstructural protein junctions and the negative regulation of protease activity to not act on other sites, and the factors affecting its conformational switching are to be answered.

This study focused on the conformational dynamics of NS2B3 protease and the role of NS2B in these conformational dynamics and cleavage site recognition. I did a thorough biophysical and biochemical characterization of NS2B3 protease and established closed conformation formation and the interface involved in its stabilization. Using in-silico molecular docking, molecular dynamic simulation, and biochemical assays, I established a possible mechanism of substrate recognition and its transfer into the catalytic pocket through substrate binding S1 pocket by NS2B.

Moreover, this study's findings gave additional information on mechanistic details of substrate recognition and its transfer into the catalytic pocket through open to closed conformation transition.

2.2 Material and Methods

2.2.1 Cloning of NS2B3, Mutants of NS2B3, NS2B48-100aa, NS3 Protease Domain, NS3CTD-NS4AN56, and GFP-Cs-Quencher Into pET24b Vector

NS2B3 of Dengue serotype-1 (Gene Bank accession number JN903579) was previously cloned into a pGEM-T vector and named pDV1-419NS2B3. DNA of NS3 protease domain coding region was amplified from (DNA coding region of N-terminal 1-172 amino acids of NS3 protein), and NS2B cytoplasmic loop (DNA coding region of 48-100 amino acids) were PCR amplified from a pDV1-419NS2B3 plasmid. These DNA fragments were linked with GGGGSGGGG linker coding DNA sequence using overlapping PCR (SM85 and SM86 used as overlapping primers) to generate a construct having 2B cytoplasmic loop (generated by PCR amplification using SM59 for and SM60), a flexible glycine serine linker, and protease domain (generated by PCR amplification using SM5 and ST1 primers) of NS3 sequentially and named NS2B3. Later this gene was inserted between the NheI and XhoI sites of the pET24b expression vector with the C-terminal His tag.

Mutant-1 of NS2B3 is a replacement of charged stretch of 'DD' at 80 and 81st position in NS2B to 'AA,' was made using overlapping PCR (primers SD1 and SD2), and the resulted gene was cloned into pET24b expression vector with C-terminal His tag. Mutant2 of NS2B3 is the replacement of conserved 'F' at 116th position NS3 to a polar residue 'R,' was made using overlapping PCR (primers SM144 and SM145) and resulted in the gene being cloned into pET24b expression vector with C-terminal His tag. Mutant-3 is a truncation of NS2B cytoplasmic region to initial ten amino acids (49-59 amino acids), and a glycine linker and protease domain comprising DNA sequence was generated using overlapping PCR and inserted in pET24b expression vector with c-terminal his tag.

Mutant-4 of NS2B3 is a replacement of a charged stretch of 'DEERDD' from 87-93aa of NS2B to 'AAAAAA,' was made using overlapping PCR (primers SM99 and SM100), and the resulting gene was cloned into pET24b expression vector with C-terminal His tag.

NS3 protease domain (1-172aa) and NS2B 48-100aa coding DNA regions were PCR amplified from pDV1-419NS2B3 vector and inserted into pET24b expression vector with C-terminal His tag independently.

NS3CTD-NS4AN56 construct was made by combining the DNA coding region of the Cterminal domain representing 475aa to 619 aa of NS3, which was PCR amplified from a pDV1-419NS2B3 plasmid with the DNA coding region of N-terminal 56aa of NS4A which was PCR (primers SM7 and SM52) amplified from pDV1-419NSP4a plasmid by overlapping PCR (SM92 and SM93) without disturbing the cleavage site and named as NS3CTD-NS4AN56 substrate. This resulted in a gene from overlapping PCR inserted into the pET24b expression vector between NheI and XhoI sites with C-terminal his tag.

GFP-Cs-Quencher construct was made by combining the DNA sequence of GFP protein with the DNA coding sequence of a decapeptide cleavage site of NS3-NS4A junction (Cs) and M2 peptide (Quencher) (**AAGRRSVSGD**CNDSSDPLVVAASIIGILHLILWILDRLLE) by two reverse primers extension PCR (primers SM147, 148, 149, and 150). This resulted in the gene having GFP protein-coding sequence towards the 5' side, the Cs site in the middle, and the M2 peptide coding sequence towards the 3'side, which was later inserted into the pET24b expression vector between NheI and XhoI sites with C-terminal His tag.

2.2.2 Expression and Purification of NS2B3, Mutants of NS2B3, NS2B48-100aa, NS3 Protease Domain, NS3CTD-NS4AN56, and GFP-Cs-Quencher Recombinant Proteins

NS2B-NS3 protease complex (from now referred to as NS2B3 for sake of brevity), its mutants, and NS2B 48-100aa recombinant proteins were expressed and purified similarly without any modification in the protocol. pET24b with NS2B3 and its mutants transformed into Rosetta DE3 cells, and cells were grown in LB Broth supplemented with 50µg ml^{-1} kanamycin and 34µg ml^{-1} chloramphenicol to an optical density at 600 nm (OD_{600}) of 0.6 to 1.0 at 37ºC. Expression of recombinant NS2B3 was induced with 0.5mM isopropyl β-D-1thiogalactopyranoside (IPTG) and incubated at 16ºC overnight; cells were pelleted by centrifugation at 5000 × g for 10 mins at 4ºC. Pellets were resuspended in a lysis buffer of 20mM Tris pH 7.4 containing 150mM NaCl, 10% glycerol, and 10mM imidazole and disrupted by sonication using a digital probe sonicator at 40% amplitude for 20 min. Both wild-type and mutant proteins are expressed as soluble proteins. The cell lysate was clarified by centrifuging at 20000 × g at 4ºC in a Sorval centrifuge with a T29 rotor for 20 min; the supernatant was taken for purification using

Ni-NTA agarose gravity column, which is pre-equilibrated with the binding buffer of 20mM Tris pH 7.4 containing 150mM NaCl, 10mM Imidazole and 10% glycerol. After the sample was loaded onto the column, the column was washed with 5 column volumes of wash buffer of 20mM Tris pH7.4 containing 150M NaCl, 50mM Imidazole, and 10% glycerol. Then bound protein was eluted with elution buffer 20mM Tris pH 7.4 containing 150mM NaCl, 300mM Imidazole, and 10% glycerol. Eluted protein was concentrated and loaded onto the Superdex 75 gel filtration column for further purification with a buffer of 50mM Tris pH9.0 containing 10% glycerol (TG buffer). A single peak of purified protein was collected, concentrated at 10-20 mg $ml^{-1,}$ and stored in TG buffer at -20ºC. For pH

studies, the purified protein was buffer exchanged into the desired buffer of different pH using Amicon ultra 10kDa cutoff concentrators according to the user protocol.

NS3 protease domain gene containing pET24b vector was transformed into Rosetta DE3 cells, and cells were grown in LB Broth supplemented with 50µg ml^{-1} kanamycin and 34µg ml^{-1} chloramphenicol to an optical density at 600 nm(OD_{600}) of 0.6 to 1.0 at 37°C. The recombinant NS3 protease domain was expressed with 1mM isopropyl β-D1thiogalactopyranoside (IPTG) and incubated at 37°C for 3 hours: cells were pelleted by centrifugation at 5000 × g for 10 mins at 4°C. Pellets were resuspended in a lysis buffer of 20mM Tris pH 7.4 containing 150mM NaCl and disrupted by sonication using a digital probe sonicator at 40% amplitude for 20 min with 15 seconds off and 5 seconds on cycles. The cell lysate was clarified by centrifuging at 20000 × g at 4°C in Sorval centrifuge with T29 rotor for 20 min, and the protein expressed as inclusion bodies were in the pellet fraction. These inclusion bodies containing pellet fraction were taken for further purification by solubilizing them in a denaturation buffer of 20mM Tris pH7.4 containing 8M urea and 150mM Nacl and purified using Ni-NTA agarose gravity column in denaturation protocol. Solubilized protein was clarified by centrifugation at 20000 × g at 4°C in Sorval centrifuge with T29 rotor for 20 min and resulted in the supernatant being loaded on NiNTA gravity column, which is pre-equilibrated with the binding buffer of 20mM Tris pH 7.4 containing 8M urea, 150mM NaCl and 20mM Imidazole. After the sample was loaded onto the column, the column was washed with 5 column volumes of wash buffer of 20mM Tris pH7.4 containing 6M Urea, 150M NaCl, 50mM Imidazole. Then bound protein was eluted with elution buffer 20mM Tris pH 7.4 containing 6M Urea, 150mM NaCl, 300mM Imidazole. Eluted protein was concentrated refolding of protein was attempted in different methods.

NS3CTD-NS4AN56 gene containing pET24b vector was transformed into Rosetta DE3 cells, and cells were grown in LB Broth supplemented with 50µg ml^{-1} kanamycin and 34µg ml^{-1} chloramphenicol to an optical density at 600 nm (OD_{600}) of 0.6 to 1.0 at 37ºC. Expression of recombinant NS3CTD-NS4AN56 was induced with 1mM isopropyl β-D1thiogalactopyranoside (IPTG) and incubated at 37ºC for 3 hours: cells were pelleted by centrifugation at 5000 × g for 10 mins at 4ºC. Pellets were resuspended in a lysis buffer of 20mM Tris pH 7.4 containing 150mM NaCl and disrupted by sonication using a digital probe sonicator at 40% amplitude for 20 min with 15 seconds off and 5 seconds on cycles. The cell lysate was clarified by centrifuging at 20000 × g at 4ºC in Sorval centrifuge with T29 rotor for 20 min, and the protein expressed as inclusion bodies were in the pellet fraction. These inclusion bodies containing pellet fraction were taken for further purification by solubilizing them in a denaturation buffer of 20mM Tris pH7.4 containing 8M urea and 150mM Nacl and purified using Ni-NTA agarose gravity column in denaturation protocol. Solubilized protein was clarified by centrifugation at 20000 × g at 4ºC in Sorval centrifuge with T29 rotor for 20 min and resulted in the supernatant being loaded on NiNTA gravity column, which is pre-equilibrated with the binding buffer of 20mM Tris pH 7.4 containing 8M urea, 150mM NaCl and 20mM Imidazole. After the sample was loaded onto the column, the column was washed with 5 column volumes of wash buffer of 20mM Tris pH7.4 containing 4M Urea, 150M NaCl, 50mM Imidazole. Then bound protein was eluted with elution buffer 20mM Tris pH 7.4 containing 2M Urea, 150mM NaCl, 300mM Imidazole, and 10% glycerol. Eluted protein was concentrated and desalted using a Hi-trap desalting column with a buffer of 50mM Tris pH9.0 containing 10% glycerol (TG buffer), and protein purified after this step was stored in -20 ºC till further use for protease assay. For pH studies, the purified protein was buffer

exchanged into the desired buffer of different pH using Amicon ultra 10kDa cutoff concentrators according to the user protocol.

A GFP-Cs-Quencher construct which was made in the pET24b vector, was transformed into BL21DE3 cells, and cells were grown in LB Broth supplemented with 50µg ml^{-1} kanamycin to an optical density at 600 nm(OD_{600}) of 0.6 to 1.0 at 37ºC. Expression of recombinant GFP-Cs-Quencher protein was induced with 1mM isopropyl β-D1thiogalactopyranoside (IPTG) and incubated at 37ºC for 3 hours; cells were pelleted by centrifugation at 5000 × g for 10 mins at 4ºC. Pellets were resuspended in a lysis buffer of 20mM Tris pH 7.4 containing 150mM NaCl and disrupted by sonication using a digital probe sonicator at 40% amplitude for 20 min with 15 seconds off and 5 seconds on cycles. GFP-Cs-Quencher protein is expressed as a soluble protein. The cell lysate was clarified by centrifuging at 20000 × g at 4ºC in a Sorval centrifuge with a T29 rotor for 20 min; the supernatant was taken for purification using Ni-NTA agarose gravity column, which is pre-equilibrated with the binding buffer of 50mM NaPO4 pH 7.8 containing 300mM NaCl, 10mM Imidazole and 10% glycerol. After the sample was loaded onto the column, the column was washed with 5 column volumes of wash buffer of 50mM NaPO4 pH 7.8 containing 300mM NaCl, 50mM Imidazole, and 10% glycerol. Then bound protein was eluted with elution buffer 50mM NaPO4 pH 7.8 containing 300mM NaCl, 300mM Imidazole, and 10% glycerol. Eluted protein was concentrated and further purified using Hitrap Q anion exchange chromatography with a buffer of 50mM Tris pH9.0 containing 100mM NaCl for binding. Bound protein was eluted using 50mM Tris pH9.0 containing 100mM to 1M NaCl linear gradient, and eluted fractions containing 32kDa size band were pooled and loaded onto the superdex-75 gel filtration column for further purification with 50mM Tris pH 9.0 containing 10% glycerol. A single peak of purified protein oligomer

eluted in the void of the column was collected and concentrated to 1mg ml^{-1} and stored in TG buffer at -20ºC.

2.2.3 Protease Assay

Protease assay was done using NS3CTD-NS4AN56 protein as substrate and NS2B3 protease and its mutants as an enzyme in a 1:5 molar ratio (protease: substrate) and incubated at 37º C for 8 hours, and samples were taken out for every 1hours time point till 4hrs after that every 2hours till 8 hours, protease assay was done in 50mM Tris H 9.0 containing 1mM CHAPS and 10% glycerol. The enzyme reaction was stopped by adding 5X SDS-PAGE loading dye and boiling at 95º C for 5mins. After completion of the assay, samples were loaded on 6-18% gradient SDS-PAGE for the separation of cleaved proteins and further analysis. To test the pH effect on the protease activity of NS2B3 protease, we used 50mM Tris pH7.0 containing 10% glycerol and 1mM CHAPS, and 50mM Tris pH9.0 containing 10% glycerol and 1mM CHAPS buffers were used for protease assay.

GFP-Cs-Quencher substrate (upon specific cleavage event at the cleavage site in between GFP protein and quencher peptide, GFP fluorescence will increase) was used in a fluorescence-based assay. This assay used GFP-Cs-Quencher at a different molar concentration (1.6uM to 6.4 uM with 8uM increment) as substrate and NS2B3 wildtype and mutant-1 of NS2B3 proteins as enzymes. Both the proteins were incubated at 37º C for 3 hours, then an increase in GFP fluorescence upon cleavage was measured in endpoint assay. For fluorescence spectra recording at 0hr time point and 3hr time point, excitation was set at 485nm, and emission scan was done from 495-535nm.

2.2.4 Size Exclusion Chromatography (SEC)

NS2B3 wild type and all of its mutants used in this study were analyzed for their quaternary structure changes at two different pHs (initial wildtype was tested three at different pHs like pH 5.0, 7.0, and 9.0, in that study at pH 5.0 five wild type protein was getting denatured and aggregated, so for later studies pH7 and pH9 buffers only used). For this analysis superdex-75 10/300 Increase GL column was used with a 0.5 ml/min flow rate. 50mM Tris pH 7.0 containing 10% glycerol and 50mM Tris pH9.0 containing 10% glycerol buffers were used for pH studies.

2.2.5 Protein Crystallization

NS2B3 protease was concentrated at 20-40mg/ml in 20mM Tris pH 7.4 containing buffer. A Grid of conditions was made from previously solved crystal structures of NS2B3 protease crystallization conditions. And drops were set in the hanging drop method.

2.2.6 Glutaraldehyde Crosslinking

Crosslinking experiment was done with two different pHs having buffers that were used with wildtype NS2B3 protease at 0.25 mg ml^{-1} concentration. For pH 7.0 buffer, 50mM Na.PO4 pH 7.2 containing 10% glycerol was used for pH 9.0 buffer, 50mM Na.CO3 pH9.4 containing 10% glycerol was used (addition glutaraldehyde changed the pH to pH7 and pH 9 to the respective buffers). Protein was incubated in the respective buffers at room temperature for 30min. Later, 1.5ul of 0.1% glutaraldehyde was added to the 28.5ul sample to make the final glutaraldehyde concentration of 0.01% in the sample and incubated for 10min at room temperature. The reaction was stopped by adding 5X SDS-PAGE loading dye and boiling at 95^{o} C for 5mins.

2.2.7 Far-UV Circular Dichroism (CD) and Intrinsic Tryptophan Fluorescence Spectroscopy

Far-UV CD and intrinsic tryptophan fluorescence spectra were recorded with NS2B3 wild-type protein at different pHs. Samples were incubated at room temperature for 30 min in the corresponding buffer before recording the spectra. Protein was used at 0.2 mg/ml concentration, and different pH buffers were used to check the pH effect on the protein secondary structure by CD spectroscopy. For pH 5.0 and 6.0, 50mM MES pH 5.0 and pH 6.0 with 10% glycerol were used. For pH 7.0- 9.0, 50mM Tris pH 7, pH8, and pH9 with 10% glycerol were used. The far-UV CD spectra were recorded using a 0.1 cm path-length quartz cuvette. Normalized mean residue ellipticity (MRE) values against wavelength spectra were calculated and plotted.

Fluorescence emission spectra were recorded using a 1 cm cuvette with an excitation wavelength set to 295 nm, and an emission scan at 300nm to 400nm was recorded. The same pH buffers were used to record intrinsic tryptophan fluorescence emission spectra to check the pH effect on structural changes in the protein.

Parameters for CD and fluorescence spectra collection and plotting were similar as explained in chapter one methods and materials (detailed in section 1.2.8).

2.2.8 FRET Assay

For the single-molecule FRET assay, two different regions were selected in NS2B and NS3 of NS2B3 protease. Cysteine and histidine were introduced by replacing threonine and isoleucine residues at 96 and 97 positions in NS2B of the NS2B3 protease construct, and another cysteine was introduced in the protease domain at the 71st position by replacing a serine, enabling labeling with a FRET fluorophore pair. After ensuring that only introduced cysteines are available for labeling, the protein was labeled with the

FRET pair, Alexa Fluor™ 488 and 594 C5 Maleimide through cysteine thiol-reactive maleimide chemistry, following protocol prescribed by the manufacturer (initial trial of sequential labeling by $CdCl_2$ protocol resulted in aggregation of protein at 1mM concentration also). Protein and fluorophore maleimide were used at a molar ratio of 1:20 in the labeling reaction. The reaction mixture in TNE buffer was incubated at RT for two hours in the dark. The labeled protein was separated from free dyes using Sephadex G25 in PD-10 desalting columns. And buffer was exchanged into 20mM Tris pH 9.0 containing 10% glycerol buffer for the assay in the presence of substrate.

Fluorescence emission spectra were recorded with an excitation wavelength set to 493 nm (λ_{max} abs. for Alexa Fluor™ 488), and emission was recorded with a wavelength range of 500-700 nm. Spectra were collected either with the labeled protein alone or with the labeled protein incubated with NS3CTD-NS4AN56 at different molar ratios for 30 min at RT.

2.2.9 Homology Modeling and Molecular Docking Simulations

Dengue NS2B3 protease of serotype-1 closed conformation structure published with PDB ID-3L6P lacks structural information from 86-100 of NS2B and Glycine linker. To add these residue coordinates and improvise the structure, we used the online homology modeling tool 'Swiss model' by taking structures of PDB ID-3L6P and PDB ID-2M9Q as templates, and the model was generated, and the same model used for molecular docking and molecular dynamic simulation. Similarly, the closed conformation homology model of NS2B3 was made using the PDB ID- 2M9Q, which has a sequence identity of more than 80% in the 'swissmodel,' and the model generated was used for molecular docking and molecular dynamic simulation studies. Decapeptide substrates were modeled

in PEP-FOLD3 software[199], representing ten residues of all six cleavage sites of NS2B3 protease in non-structural proteins and Capsid-prM.

For molecular docking of all decapeptide substrates onto the NS2B3 protease, decapeptide substrates were taken as ligands, and NS2B3 protease, both open and closed conformations were taken as receptors in the online 'Cluspro' docking tool[176]. Interactions of the balanced format of output were taken for analysis.

Simulated docking was done with a closed conformation homology model of NS2B3 protease with NS3- NS4A cleavage site representing deca-peptide substrate in DockThor server[200]. In DockThor server, on the closed conformation, local docking was by specifying grid for the dock (including the catalytic pocket).

2.2.10 Molecular Dynamic Simulations

MD simulation experiment was done to study the 'basic' pH effect on NS2B3 protease conformation compared to neutral pH. The homology model of NS2B3 protease open conformation is used as an input at the PlayMolecule[201] server and is available at https://playmolecule.org/ for assignment of protonation states at a given pH value (pH 7.0 and pH 9.0). Protein was solvated with water (water model-TIP3P), and no ions were added. The system was energy minimized in the NVT ensemble and equilibrated for 125 picoseconds in the NPT ensemble. Each system was simulated using Gromacs-5.1 with CHARMM36 forcefield with a 1fs step time. Periodic boundary conditions were applied in the simulation. Following equilibration, the system was subjected to simulation production for 10ns at 300K. Similarly, the NS3-4A cleavage site representing the decapeptide model docked structure of NS2B3 protease with open conformation at pH 9.0 was taken for the next MD simulation experiment in identical conditions for 20ns at 300K.

2.3 Results

2.3.1 NS2B3 Protease Is More Active at pH 9.0

The recombinant NS3protease-NS2B complex (either expressed as a single protein connected by flexible Gly-Ser linker or as individual proteins) is used for by biophysical, biochemical characterization of the interaction and the protease activity.

The independent expression of NS2B 48-100aa region and NS3 protease domain resulted in NS2B in the soluble fraction and NS3 protease expressed in an insoluble form. The NS3 protease was purified from inclusion bodies through Ni-NTA chromatography in the denaturation method (Figure 2-3B). Refolding of the NS3 protease domain resulted in aggregation, indicating it needs NS2B for its proper fold, which was reported previously[87]. For the NS2B3 protease construct, we used the well-established approach of linking the NS2B cofactor region with the protease domain coding region of NS3 with glycine serine linker without any further modification, resulting in a construct named NS2B3 protease (Figure 2-3A).

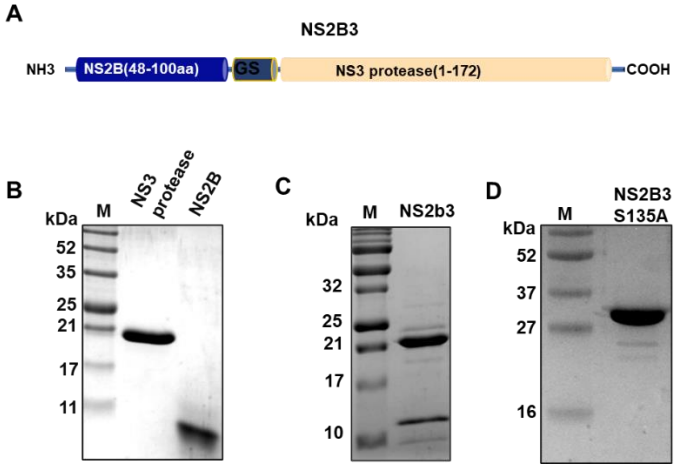


Figure 2-3 NS2B3 expression and purification (A) Schematic of NS2B3 construct (B) SDSPAGE of NS2B and NS3 protease purified recombinant proteins (C) SDS-PAGE of NS2B3 protease recombinant purified protein (D) SDS-PAGE of NS2B3 active site mutant

NS2B3 protease was expressed in a soluble fraction from Rosetta DE3 cells and purified in native conditions (Figure 2-3C) by gel filtration chromatography followed byNi-NTA

chromatography. The recombinant NS2B3 protease was purified from Rosetta-DE3 cells and showed two bands in SDS-PAGE. Furthermore, this purified NS2B3 protease protein with two fragments was eluted as a single peak in SEC. These two bands' corresponding sizes match the independent protease domain, and the NS2B cofactor resulted from a cleavage event at the junction after the glycine linker, which was reported previously[202] for similar constructs. The inactive catalytic serine mutant protein of the same construct did not show this cleaved band, confirming the autocleavage evet at the glycine liker site of active protein (Figure 2-3D). So I proceeded with this protein for biophysical and biochemical assays.

For functional assessment of the protease (for protease assay), a substrate construct was made by combining NS3 C-terminal domain (476-619) with NS4A N-terminal 56 residues using overlapping PCR (Figure 2-4A), resulting in a protein expressed with a canonical cleavage site of NS2B3 protease mimicking the original polyprotein, which was later named NS3CTD-NS4AN56.

NS3CTD-NS4AN56 protein was expressed and purified from RosettaDE3 cells from inclusion bodies using the denaturation protocol with 6M urea. Purified protein was slowly refolded using the on-column refolding method on the Sephadex G-25 desalting column (Figure 2-4B) by removing the denaturant. This purified NS3CTD-NS4A protein was buffer exchanged into the corresponding buffer for protease assay.

The protease assay with NS3CTD-NS4AN56 substrate will give two fragments after cleavage at the scissile bond by NS2B3 protease corresponding to ~17kDa and ~7kda size band on SDS-PAGE (Figure 2-5A).

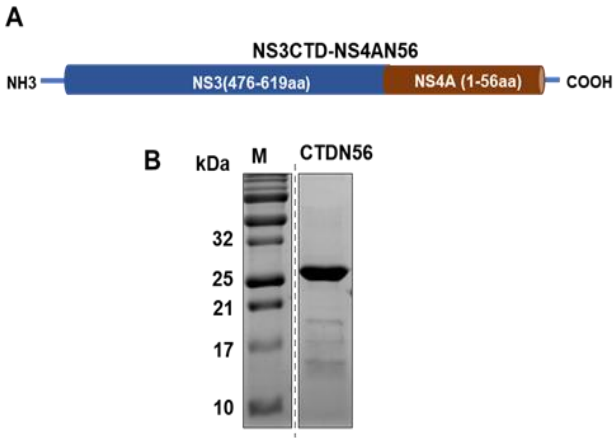


Figure 2-4 NS3CTD-NS4AN56 expression and purification (A) Schematic of NS3CTDNS4AN56 construct (B) SDS-PAGE of purified NS3CTD-NS4AN56 recombinant protein

Previous literature showed that NS2B3 protease activity would be more at pH 8.5 and pH 9.0. So initially, the protease assay was done at pH 9.0 and 7.0; NS2B3 protease showed better/maximum activity at pH 9.0 than at pH 7.0. For further functional assays with mutants, the protease assay was optimized with a 1:5 molar protease ratio of the protease

(NS2B3 protease) to the substrate (NS3CTD-NS4AN56) at pH 9.0 with 10% glycerol and 1mM CHAPS. The activity of the NS2B3 protease was assessed by the intensity of two cleavage products on 6-18% gradient SDS-PAGE at their respective molecular weights ~17kDa, and ~7kDa after a specific cleavage event at the junction of NS3-NS4A.

Purified NS2B3 protease is active, and the activity increased for two hours; later, it showed no increase and stayed the same till 8 hours. NS2B3 showed slower kinetics, and only two specific cleavage bands of NS3CTD(~17kDa) and NS4AN56(~7kDa) were increased upon incubation for different time points (Figure 2-5B).

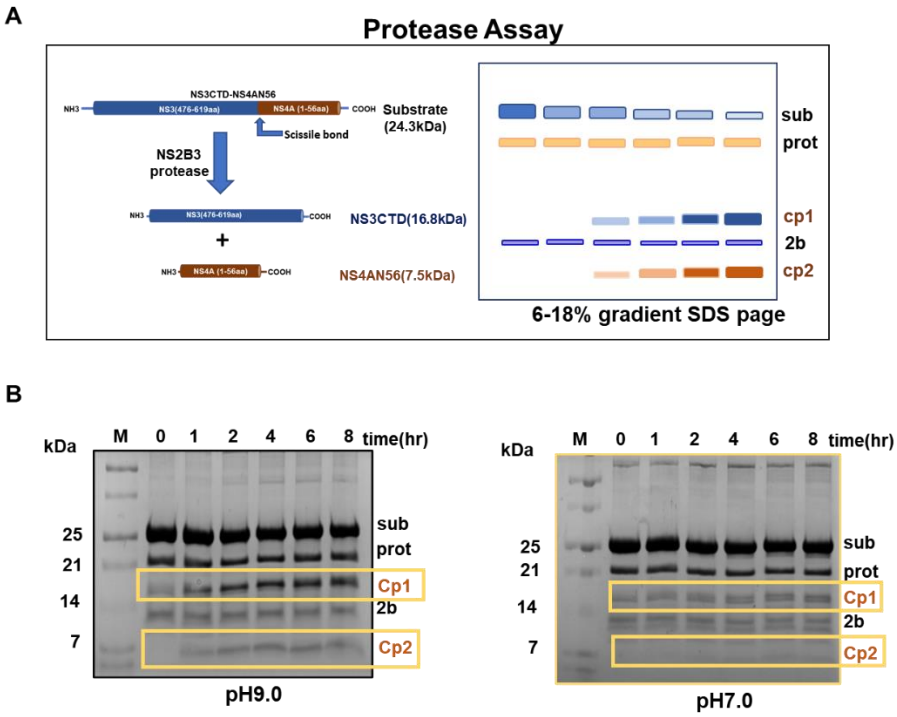


Figure 2-5 Protease assay with wildtype NS2B3 protease (A) Schematic of protease and predicted cleavage bands (B) SDS-PAGE (6-18% gradient) analysis of NS2B3 protease assay with NS3CTD-NS4AN56 substrate at pH 9.0 and pH 7.0

A Series of biophysical experiments were conducted on NS2B3 protease to test whether any structural changes in NS2B3 at this 'basic' pH (pH 9.0) contributed to the increased protease activity.

Secondary structure analysis by Far-UV circular dichroism spectroscopy shows a typical and unique spectral signature like a chymotrypsin fold with NS2B3 protease at neutral pH and basic pH [203] (Figure 2- 6A). NS3 protease lost significant secondary structural content below pH 6.0, and we also observed visible denatured and aggregated protein at this low pH. Above pH 6.0, there are no significant structural changes; CD spectra at pH 7.0 and pH 9.0 are overlapped with each other (Figure 2-6A).

The intrinsic tryptophan fluorescence spectroscopy experiment of NS2B3 protease at three different pHs showed no significant change in the emission spectra (Figure 2-6B).

Both these experiments confirm no significant structural change in the protein upon pH change.

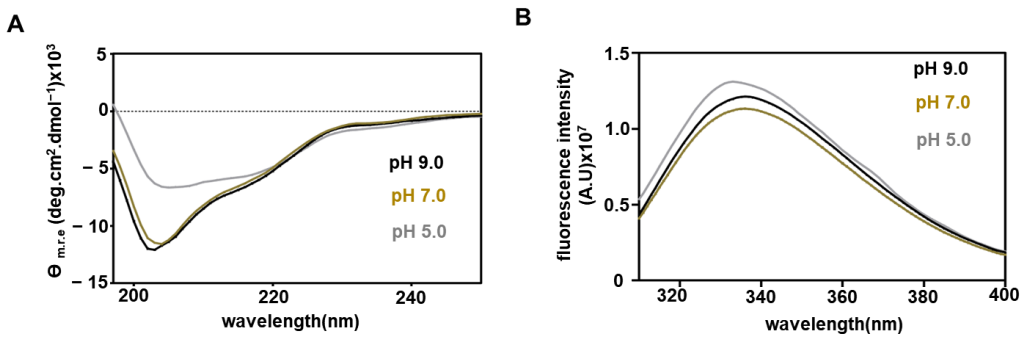


Figure 2-6 pH effects on the structure of NS2B3 protease (A) CD spectra of NS2B3 protease at pH 5.0, pH7.0, and pH 9.0 (B) Intrinsic tryptophan fluorescence spectra of NS2B3 protease at pH 5.0, pH7.0, and pH 9.0

In size exclusion chromatography (SEC) of NS2B3 protease, at different pHs (including pH 9.0, pH 7.0, and pH 5.0) eluted at different retention volumes, eluted at 9.32ml retention volume in pH 9.0 buffer, at 11.2ml retention volume in pH 7.0, and at 12.5ml retention volume in pH 5.0 (most of the protein got aggregated, indicating protein is unstable at this low pH 5.0). These retention volumes correspond to the different oligomeric forms of NS2B3 protease (Figure 2-7A) at pH 9.0 has a tetramer form of protein (runs at around ~98kDa globular protein size), at pH 7.0 has a dimer form (runs at around ~44kDa globular protein size). NS2B co-factor also showed similar oligomeric behavior with pH (Figure 2-7B).

The previous literature where serotype-1 NS2B3 crystal structure (PDB-3L6P) solved showed two different biological assemblies corresponding to dimer and tetramer forms picked by PISA. Structural analysis of these two biological assemblies showed that the dimer form has the interface shared through NS3-NS3 inverse orientation interaction (Figure 2-7C), hiding the catalytic pocket, and the tetramer form mediated by 2B cofactor

conserved beta-hairpin region with hydrophobic interaction and the catalytic pocket is solvent accessible (Figure2-7D).

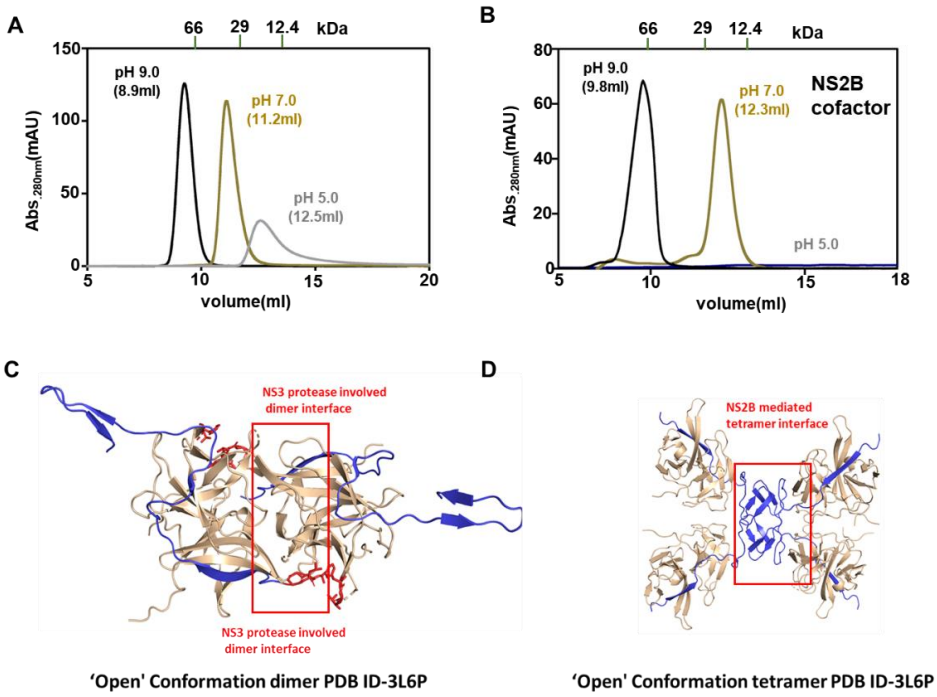


Figure 2-7 NS2B3 pH mediated oligomeric forms (A) Size exclusion chromatogram of NS2B3 protease at pH 5.0 (grey), pH 7.0 (gold) and pH 9.0 (black) (B) Size exclusion chromatogram of NS2B cofactor at pH 5.0 (grey), pH 7.0 (gold) and pH 9.0 (black) (C) Dimer form NS2B3 (PDB ID-3L6P) (D) Tetramer form of NS2B3 (PDB ID-3L6P)

Glutaraldehyde crosslinking-based stabilization assay was performed to validate these oligomeric conformations of NS2B3. In tetramer of NS2B3 held by hydrophobic interaction, the interface does not have any primary amine group having residues and in this catalytic pocket has open access (2-7D), after glutaraldehyde treatment upon loading onto the SDS-PAGE will give monomer band predominantly. The dimer conformation interface has primary amine having lysine residue in the interface, which can be crosslinked to stabilize the dimer form and block the catalytic pocket partially. The glutaraldehyde crosslinking assay at pH 7.0 and 9.0 showed NS2B3 at pH 9.0 is predominantly monomer in SDS-PAGE, whereas pH 7.0 showed a decrease in monomer

and an increase in dimer band (Figure 2-8). Indicating its conformational diversity and further confirming its oligomeric behaviour and why protein is more active at pH 9.0.

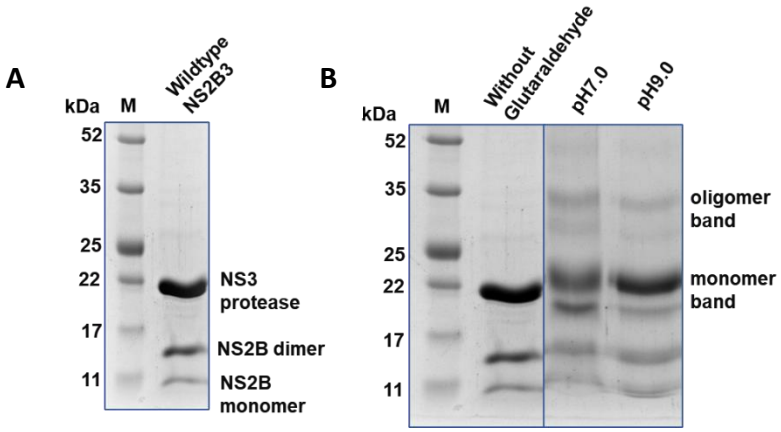


Figure 2-8 Glutaraldehyde crosslinking assay with NS2B3 protease at pH 7.0 and pH 9.0 (A) Control SDS-PAGE of NS2B3 without glutaraldehyde, (B) After glutaraldehyde addition at two different pHs.

2.3.2 Conformational Change of NS2B3 from 'Open' to 'Closed'

There are multiple conformations for NS2B3 protease of DENV, although the substrate mimicking inhibitor bound form is always monomeric (ex: PDB ID-2M9Q) and in the closed conformation (Figure 2-8). It was shown that NS2B3 would fluctuate between open and closed conformation through NMR and MD simulation studies[204],

The crystal structure of NS2B3 protease of DENV-1 (PDB ID-3L6P) has lost some of the structural information in the NS2B cofactor region and does not have closed conformation. Crystallization of NS2B3 protease is attempted to fill this structural gap and get a closed conformation crystal structure for DENV-1 NS2B3 protease. However, in some crystalization conditions, microcrystals for NS2B3 protease are formed after 'three' months but attempts to get diffraction quality crystals failed. A closed conformation structure and open conformation structure homology models were

generated in swissmodel (Figure 2-9) to fill the gaps in the structure and get a closed conformation model for DENV-1 NS2B3 by taking PDB-ID 2M9Q and 3L6P as templates.

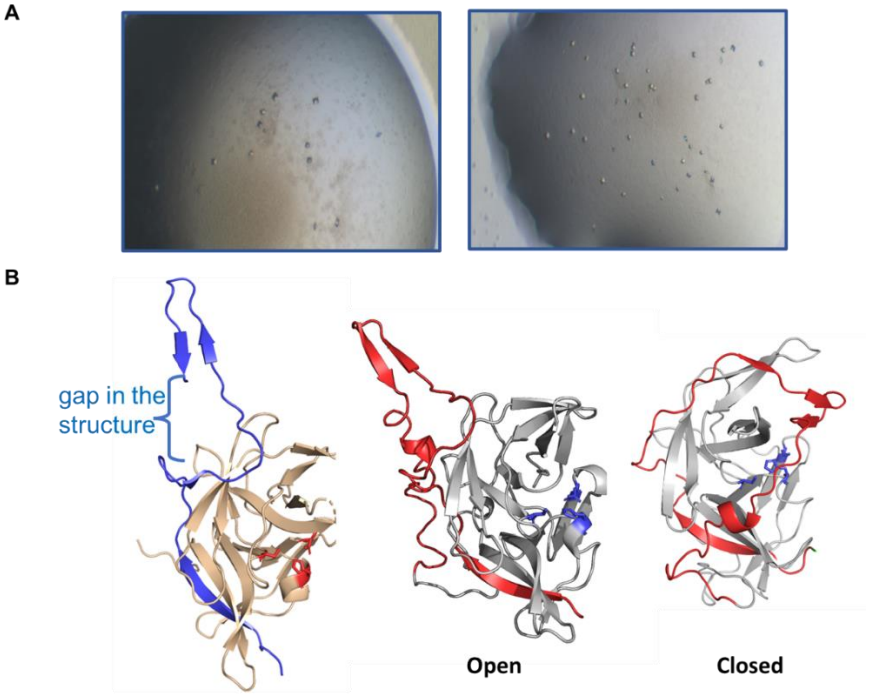


Figure 2-9 Crystallization of NS2B3 protease and Homology models of DENV-1 NS2B3 (A)NS2B3 protease crystals (B) NS2B3 open and closed conformation homology models generated in 'swissmodel'

This improvised version of (PDB-ID 3L6P based) DENV-1 open conformation with dimer form of NS2B3 protease homology model was prepared for constant pH simulation in 'playmolecule' server (https://www.playmolecule.org/) at pH 7.0 and pH 9.0. With CHARMM36 forcefield in GROMACS software, 10ns simulation was performed without any ions.

After the simulation of 10ns, RMSD of NS2B3 backbone at pH 9.0 was more compared to pH 7.0 (Figure 2-10A), indicating protein is more dynamic at pH 9.0. RMSF of NS2B cofactor region showed up to ~1.4 nm (Figure 2-10B), specifically in a structurally

conserved beta-hairpin motif (73-86amino acids region) of NS2B, which will be positioned differently in the closed conformation. The trajectory file showed NS2B cofactor beta-hairpin motif dynamic motion towards NS3 protein catalytic pocket, where it interacts in the closed conformation (Figure 2-10C). These results show how pH 9.0 induces or accelerates the structural changes in protein towards the formation of closed conformation but may not stabilize closed conformation (two new interfaces that will emerge in the closed conformation will stabilize the closed conformation), and this stabilization may require substrate binding.

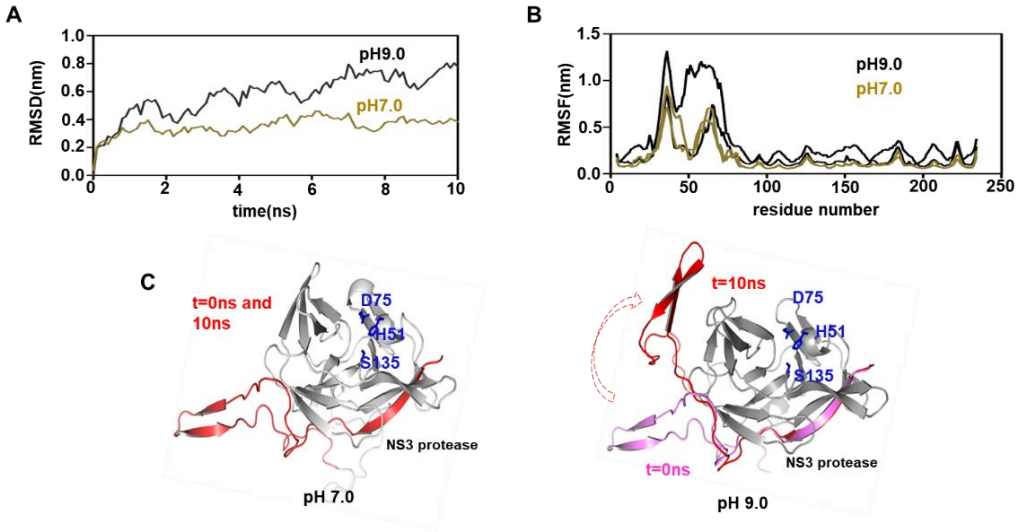


Figure 2-10 Molecular Dynamics Simulation of NS2B3 open structure (A) Backbone RMSD analysis of NS2b3 at pH7.0 (gold) and pH9.0 (black) (B) RMSF analysis of NS2b3 at pH7.0 (gold) and pH9.0 (black) (C)Conformational snapshots at t=0ns and t=10ns at pH 9.0 (motion in conserved structural motif movement towards catalytic pocket showed in red color dotted arrow)

2.3.3 Residues Involved in Stabilization of 'Closed' conformation

MD simulation of open conformation showed dynamics of the protein towards closed conformation. The structure analysis of both open and closed conformations in MOLEonline (https://mole.upol.cz/) software showed tunnels that can access the substrate to bind and get cleaved in both open and closed conformations (Figure 2-11).

The tunnel near the catalytic pocket in open conformation is accessible due to it is not covered by any other secondary structures. However, a closed conformation catalytic pocket is covered by an extension of NS2B. The surface APBS calculation of the tunnel formed near the catalytic pocket in closed conformation becomes electronegative. And the two new interfaces that emerged in the closed conformation may stabilize the structure in the substrate-bound form to hold a substrate (K/R at P1or P2 positions) through ionic interaction for cleavage.

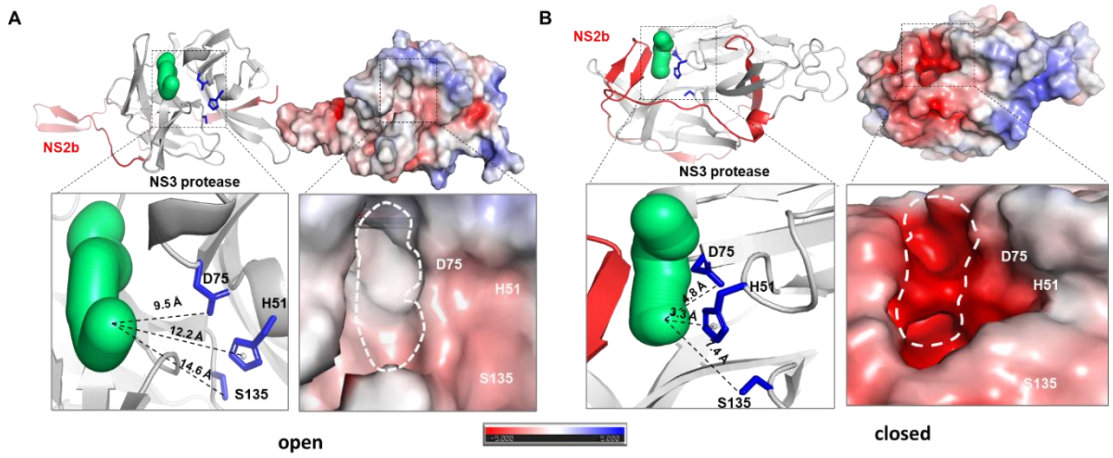


Figure 2-11 Tunnel analysis of NS2B3 'open' and 'closed' conformations (A) Open conformation, tunnel (green surface/sphere representation) formation near catalytic pocket, with inset zoomed image, right- electrostatic surface representation of catalytic pocket with catalytic triad marking, (B) Closed conformation, tunnel (green surface/sphere representation) formation near catalytic pocket, with inset zoomed image, right- electrostatic surface representation of catalytic pocket with catalytic triad marking, with inset zoomed image.

Two new interfaces will be formed between the NS2B cofactor and NS3 protease in the closed conformation (Figure 2-12A), and critical residues involved in these interfaces are found through structural analysis and multiple sequence alignment. Multiple sequence analyses of NS3 showed a conserved ‘F118’ residue along with ‘T’ and ‘P’ in the first interface, and the second interface will be formed by ‘80DD81’ in the beta-hairpin motif of NS2B with the ‘73KK74’ of NS3 (Figure 2-12B).

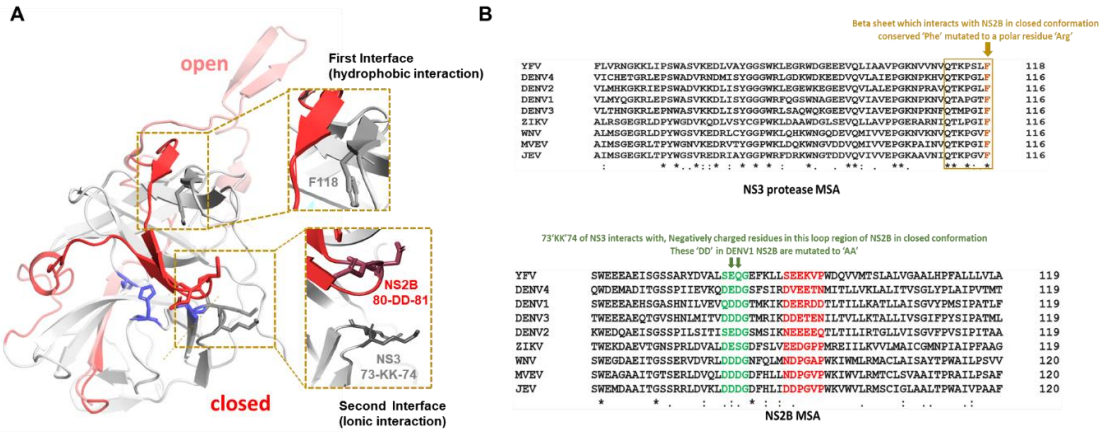


Figure 2-12 Structure analysis of closed conformation and MSA of NS2B and NS3 (A) Cartoon representation of NS2B3 closed conformation with two interfaces (zoomed out) involved in its stabilization (B) MSA of NS2b and NS3 sequences

Two independent constructs were made with these mutations and named mutant-1(in NS2B at 80DD81 to AA) and mutant-2 (NS3 F118 to R) to disturb these interfaces and destabilize the closed conformation (Figure 2-13A).

In the protease assay with mutant-1, very faint cleavage bands were observed on SDSPAGE after four hours, where specific cleavage products will appear compared to WT NS2B3 protease. Similarly, protease assay with mutant-2 also showed very faint cleavage bands weaker than in mutant-1(Figure 2-13B), indicating the drastic reduction in the

activity.

With the existing structural data and these results, it is evident that these two regions are involved in closed conformation stabilization, and the formation of a closed conformation is absolutely essential for NS2B3 protease activity. However, the dynamic motion which leads to the open to close conformation should be validated experimentally.

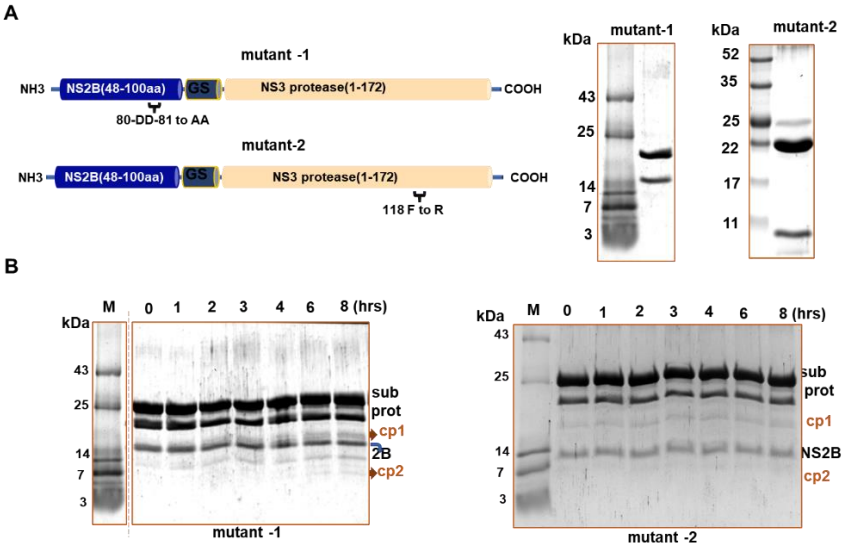


Figure 2-13 Mutational probing of NS2B3 to disturb close conformation (A) Schematic and SDSPAGE of purified recombinant proteins of mutant-1 and mutant-2 (B) SDS-PAGE of NS2B3 protease mutant-1 and mutant-2 protease assay

A FRET (fluorescence energy resonance transfer) experiment was performed to validate the structural dynamics of closed conformation formation in the presence of substrate. For the FRET assay, two positions were chosen, one in NS2B and the other in NS3 (Figure 2-14A), and these two positions were mutated into cysteine and resulted in a mutant protein labelled with fluorophores FRET pair, Alexa Fluor™ 488 and 594 C5 Maleimides. FRET assay was conducted by titrating the unlabelled substrate against NS2B3 protease at pH 9.0. The R_o of the FRET pair, Alexa Fluor™ 488 and 594 is bigger than the distance between the labelled residues, so to analyze the FRET analyzed, the donorem519/acceptor em617 ratio was taken as a measure of FRET to validate the dynamics. With titration substrate of NS2B3 protease, the em519/em617 ratio decreases compared to the buffer control (Figure 2-14B), indicating the dynamic motion of domains in NS2B3 protease possibly NS2B motion closer towards NS3 protease forming a closed conformation.

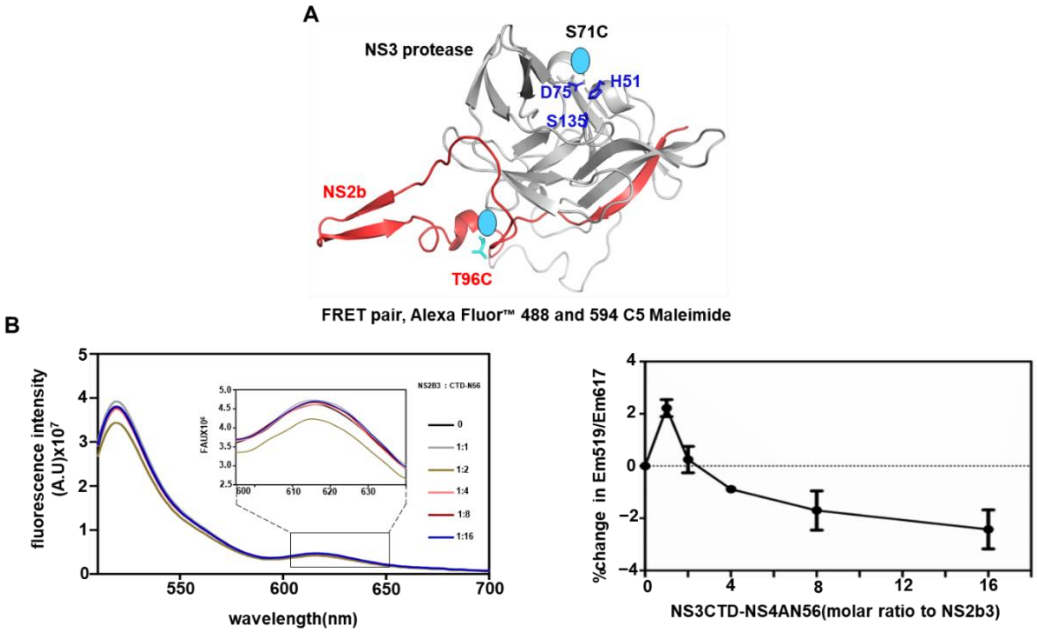


Figure 2-14 FRET assay of NS2B3 in presence of substrate (A) labeling of NS2B3 with FRET pair (B) FRET assay of NS2B3 titrated with the substrate at pH 9.0

Both FRET and MD simulation results indicated pH 9.0 induces conformational changes towards open to closed conformation, and the closed conformation may be stabilized by substrate binding in vitro.

2.3.4 NS2B Is Involved in Protease Substrate Specificity

In DENV serotype-1 polyprotein, 18 sites have double basic amino acids and small residue immediately next to them, which may mimic cleavage sites of NS3 and cause abortive poly-protein processing. However, NS3 specifically recognizes only the cleavage site in the cleavage junction between two proteins where it has to be cut. So, we analyzed all these sequences; all cleavage site junctions of NS3 are either in loops or in coils giving clues about its accessibility. However, the residues involved in NS2B3 protease P1 site recognition (S1 pocket) were identified in DENV, but the residues

involved in the recognition of second-basic residue fluctuating between P2 to P5 and how this cleavage site is transferred into the catalytic pocket is still a question.

In the previous tunnel analysis of closed conformation, the catalytic pocket was surrounded by negatively charged (87-93) residues of NS2B in the closed conformation (Figure 2-10A), indicating that this electronegative pocket may be involved in substrate recognition.

To test the role of this electronegative pocket in substrate recognition, modeled decapeptide substrates representing all cleavage sites in PEPFOLD3 were used as substrates for docking in the 'cluspro' online supercomputer as ligands and both closed and open conformations as receptors for a random rigid body global docking. This resulted in balanced interactions having dock poses with the largest cluster size, and low energy scores were taken for analysis. All the dockings of peptides gave one or clusters that are almost similar, and in these clusters, all the peptides docked into this electronegative pocket created by negatively charged ranging from 87-93 residues (DEERDD) of NS2B in both open and closed conformation (Figure 2-15A).

The simulated local docking of the NS3-NS4A cleavage site representing decapeptide with the closed conformation of NS2B3 protease in the 'DockThor' server showed that the tunnel lining the S1 pocket near the catalytic triad got narrowed by sidechain orientation change of S1 pocket lining Y161, Y150 after substrate binding. And these S1 pocket lining residues moved towards the P1 'arg' side chain to make possible 'pi-cation' interaction to be held by this pocket may guide the scissile bond between P1 'arg' and P1' ser to the catalytic pocket (Figure 2-15B). In both rigid-body docking and simulated docking, P2 'arg' of the NS3-NS4A cleavage site was held by the electro-negative pocket of NS2B.

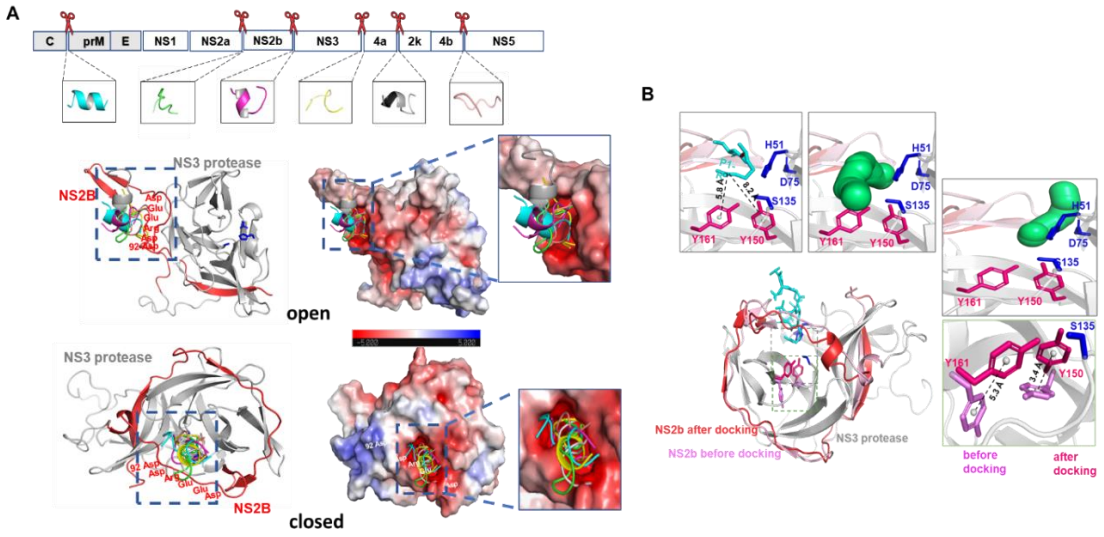


Figure 2-15 Docking of decapeptide substrates with NS2B3 (A) Cluspro global docking with decapeptide as ligand and NS2B3 open and closed conformations as receptor monomer docked pose with all substrates (zoomed out image of APBS electrostatic surface representation with peptide substrate docked onto electronegative pocket) (B) Simulated docking of Closed conformation of NS2B3 with NS3-NS4A representing decapeptide (zoomed out image of NS2B3 S1 pocket with tunnel and residues lining S1 pocket side chain orientation before and after docking)

Two mutants, mutant-3 and mutant-4, were created to validate this negatively charged 87-93 residues stretch of NS2B role in substrate specificity and recognition. Mutant-3 is a 48 to 59aa amino acids stretch of NS2B linked to the protease domain of NS3 with a glycine linker, and the rest of the NS2B cofactor region was deleted (Figure 2-16A). In mutant-4, the negatively charged 87-93 residues region of NS2B is replaced with a multiple alanine stretch (‘AAAAAA’) which makes the substrate recognition pocket bigger, uncharged, and allows any peptide into the catalytic pocket (Figure 2-16B). Both mutant-3 and 4 are expressed and purified as soluble proteins in native conditions. A protease assay was performed using these mutant proteins as enzymes and NS3CTDNS4AN56 protein as substrates at pH 9.0 to see the effect of the mutation. Both the mutants showed non-specific multiple cleavage bands (Figure 2-16C), which increased through the incubation

period in SDS-PAGE, indicating protease lost its specificity towards the cleavage site and became non-specific. These results confirmed that these negatively charged 87-93 residues of NS2B are essential for substrate specificity, recognition, and regulation of protease activity.

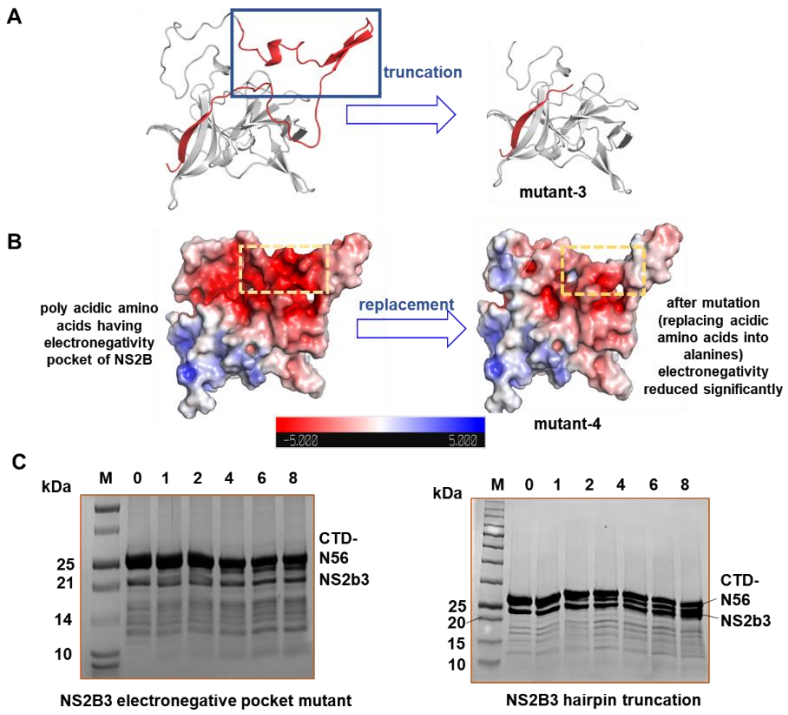


Figure 2-16 Electronegative pocket of NS2B3 mutants (A) mutant-3 cartoon representation (B) mutant-4 electrostatic surface representation (C) SDS-PAGE of mutant-3 and mutant-4 protease assay

After substrate recognition in open conformation at NS2B electronegative pocket, how it reaches the catalytic pocket reaches is the question of concern. To answer this question, an MD simulation experiment was performed at constant pH 9.0 with NS2B3 protease docked with NS3-NS4A cleavage site representing decapeptide in the open conformation. The simulation was conducted for 50ns in Gromacs with Charm33 forcefield. The RMSD of the NS2B cofactor is more (Figure 2-17A), and then in the beta-hairpin motif and electronegative pocket, RMSF is more (Figure 2-17B). The cluster analysis with a 5Å

cutoff gave three ‘major’ clusters; cluster-3 representation is below ten which is negligible. Comparative analysis of cluster-1 and cluster-2 showed that the distance between P1’arg’ sidechain and the S1 pocket representing ASP129 was reduced by 5.2 Å in cluster-2 (Figure 2-17C).

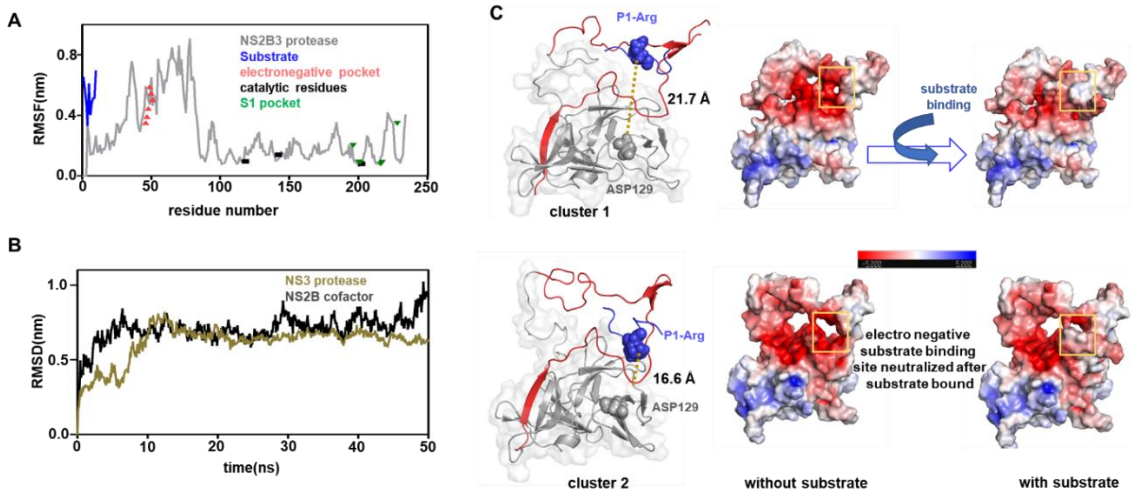


Figure 2-17 MD simulation of NS2B3 protease with NS3-NS4A cleavage site representing decapeptide (A) Backbone RMSF analysis of NS23 in the presence of substrate (B) RMSD analysis of NS2B3 in the presence of substrate, (C) Cluster analysis of 50ns simulation top- cluster 1 cartoon and APBS electrostatic surface representation with substrate P1 ‘arg’ and S1 pocket representing ASP129 in sphere representation, down- cluster 2 cartoon and APBS electrostatic surface representation with substrate P1 ‘arg’ and S1 pocket representing ASP129 in sphere representation

The substrate-bound near the electronegative pocket of NS2B is moving along with the substrate towards the S1 pocket of NS3 protease; this S1 will hold the side chain P1’arg’ during scissile bond cleavage and may help in orienting scissile bond towards catalytic serine. In the NS3-4A decapeptide, P2 ‘arg‘was held by NS2B while P1 ‘arg’ was in NS3’s S1 pocket, giving additional specificity to recognize the cleavage site.

2.4 Discussion

NS3 protein is the center of the replication complex established by several studies due to its interaction with the majority of the nonstructural proteins. The two-component NS2B3

protease complex is critical for the replication complex and vesical replication formation. Most functional nonstructural protein formation from polyproteins is done by NS2B3 protease-mediated polyprotein processing. So, the first steps lead to the formation replication complex starting from the NS2B-NS3 interaction. The remaining nonstructural protein interactions may happen during or after polyprotein processing. NS2B-NS3 interaction may be a co-translational event. The functional importance of NS2B-NS3 interaction is for polyprotein processing, established. The structure of NS2B3 protease has two major conformations: a beta strand ranging from 45-56 aa region sandwiched between the beta-sheet of NS3 protease. However, the later part of the NS2B cofactor (57-100aa) has some conformational flexibility and is responsible for these two different conformations of NS2B3 protease, and mutations in this region showed it is critical for the NS2B3 protease activity.

The cleavage site of NS2B3 has some sequence conservation with two or more basic residues towards the N-terminal of the scissile bond with the P1 position always occupied by R/K. Although similar patterns having sequences are present throughout the polyprotein (21 sites in DENV1 polyprotein) with multiple ‘basic’ residues followed by a small side chain having residue, most of these cleavage sites are not accessible (which are buried inside the structure) for the NS2B3 protease. The sequence pattern of the cleavage site and its accessibility will contribute to the NS2B3 protein cleavage specificity at the junctions between nonstructural proteins. However, the mode of action, how NS2B3 protease recognizes substrates (cleavage sites), the structural changes it undergoes from open to closed conformations, and how these changes are induced are the questions we tried to answer in this chapter.

Glycine serine linker having recombinant NS2B3 protease expressed and purified from RosettaDE3 auto-cleaved at the glycine linker junction to give NS2B cofactor and NS3 protease, which are already interacted and co-eluted together from SEC. The initial biophysical and biochemical characterization of NS2B3 protease showed it would form a dynamic oligomer based on pH mediated by NS2B. Protease assay with NS3CTDNS4AN56 substrate at pH 9.0, NS2B3 protease showed more activity than at neutral pH

(pH 7.0) which is evident from two specific cleavage bands on intensity at 17kDa (NS3CTD) and 7kDa (NS4AN56) on SDS-PAGE is more in pH 9.0 compared to pH 7.0. The glutaraldehyde crosslinking study confirmed the formation of dimer at pH 7.0 and given hints about the hydrophobic interaction-based tetramer conformation.

From the mutational and structural studies, it was believed that closed confirmation is the active conformation of NS2B3. Moreover, recent MD simulation studies showed that NS2B3 protein fluctuates between open and closed conformation in a μs simulation which is very slow for an enzyme [204]. To test if pH 9.0 is accelerating this conformational change or not, we did a constant pH MD simulation on open conformation without any substrate. Later the MD simulation studies at pH 9.0 confirmed that NS2B3 protease is more dynamic. Moreover, from the trajectory file, the motion of the conserved structural betahairpin motif is towards the catalytic pocket; from these results, it is evident that at pH 9.0, NS2B3 protease has a high tendency to form a closed conformation.

The major structural difference between open to closed conformation is that two more new interaction interfaces will emerge in the NS2B3 structure, which may stabilize the closed conformation. To find the residues involved in this closed conformation stabilization, structures of previously solved closed conformation were analyzed. The

region of 57-100aa of NS2B is responsible for the closed conformation formation, and mutations in this region reduce the activity of NS2B3 protease. This same extended region will interact with NS3 in closed conformation by forming two new interfaces of interactions. However, residues that are critical in this interaction from the NS3 side were not established; in this study, a conserved phenylalanine residue in a hydrophobic patch in the first interaction interface was mutated to a polar Arg to disturb hydrophobic interaction essential for closed conformation stabilization.

Along with this, a double aspartate in the DENV serotype-1 loop region structurally conserved bet-hairpin motif (73-86amino acids region) of NS2B came very close to 73‘KK’74 of NS3 in substrate-bound closed conformation of NS2B3 protease forming a second interface may also be involved in stabilization of closed conformation[193,197]. Mutants that will disturb these interactions were made and tested that the closed conformation stabilization is necessary or not for the protease activity. Both mutant proteins showed a drastic reduction in the protease activity, confirmed these two newly emerged interaction interfaces of NS2B3 protease closed conformation is critical for protease activity.

However, these dynamics open to closed conformation formation must be validated experimentally. A single molecule FRET experiment was performed by labeling NS2B and NS3 protease of NS2B3 protease, which are positioned away from each other in the open conformation, and come close in the closed conformation. The results from the FRET experiment, in line with MD simulation data, further confirmed motion in NS2B when the substrate-bound at pH 9.0. Along with this in-silico experimental data, substrate-bound NS2B3 closed conformation crystal structures and FRET assay indicated that NS2B3 will fluctuate between open-to-closed conformation accelerated by ‘basic’

pH and open-to-closed conformation stabilization by an attained by the bound substrate in the catalytic pocket.

In the closed conformation, the catalytic pocket was covered by part of the NS2B cofactor and created an electronegative pocket near the catalytic pocket; the majority of this electronegativity was contributed by NS2B (87-93aa region with multiple aspartates). The cleavage site with ‘basic’ residues will be electropositive, which may be recognized by this electronegative pocket of NS2B.

In the molecular docking simulation studies, the modeled decapeptide substrates representing all six cleavge sites NS2B3 protease were docked onto this electronegative pocket of NS2B (87-93aa region with multiple aspartates) with the largest cluster size of docked poses and lowest energy score. This electronegative pocket is away from the catalytic pocket in the open conformation and laid onto the catalytic pocket in the closed conformation. The interaction analysis of these docked poses showed that the cleavage site's second ‘basic’ residue (other than P1) interacted with one of the acidic residues of the electronegative pocket of NS2B. Mutational replacement of this electronegative pocket of NS2B with alanine strecth to neutratlize and subsequent protease assay with mutant protein showed multiple cleavage bands (more than four) with NS3CTDNS4AN56 substrate increased with time instead of two specific cleavage bands. Similar results were obtained when a truncated NS2B cofactor region (57-100aa region deletion) having NS2B3 protease was used as an enzyme in the protease assay.

Chymotrypsin fold having serine proteases like trypsin and chymotrypsin surface loops played a major role in substrate recognition. Here in the case of DENV NS2B3 protease, the surface's electronegative loop region of NS2B also played a critical role in the recognition. The MD simulation analysis of NS3-NS4A cleavage site representing

decatpeptide substarte bound NS2B3 protease docked structure gave hints towards the possible path of substrate reaching the catalytic pocket.

A schematic model as a probable explanation of NS2B cofactor determining the substrate specificity for NS3 protease is presented below:

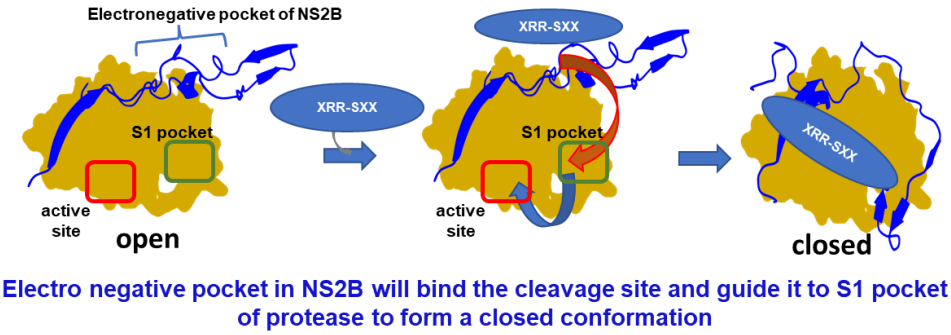


Figure 2-18 Mechanistic model of NS2B3 protease substrate recognition and open to closed conformation transition

As per this model that we proposed here, an electronegative pocket generated on the NS2B3 surface due to acidic residues stretch (DEERDD) in NS2B is probably the site where the substrate binds to. The electronegative stretch provides a charge complimentary site for the XRR-SXX motif seen in the NS3protease cleavage sites in the polyprotein. A conformational switch of the substrate-bound NS2B from open to close state may bring the substrate along close to the catalytic pocket, handing over the side chain of P1 ‘arg’ into the S1 pocket and holding the ‘basic’ residue in the cleavage site.

Chapter 3: Biochemical Characterization of DENV NS4A Protein and Its Interaction With NS3 Protein

3.1 Introduction

Among the four transmembrane NS proteins of DENV, NS2A and NS4B are involved in anchoring other soluble proteins inside replication vesicles. Besides anchoring the replication complex inside the vesicle, these membrane proteins have distinct functions in viral replication. NS2A is essential for replication virus particle synthesis and was shown to colocalize with other structural proteins and NS3[205]. NS4B was shown to interact with NS3 and helps enhance its helicase activity[149]. NS2B acts as a cofactor for NS3, needed for proper processing of polyprotein, and it is observed that NS2B bound NS3 loses its affinity towards DNA and prefers only RNA as a substrate for its unwinding activity[97]. However, NS4A is critical for viral replication, and the exact mechanistic role of NS4A in replication needs to be deduced further.

Evidence for colocalization of NS4A and other NS proteins involved in replication is available from electron tomography and confocal microscopy colocalization studies with other nonstructural proteins and dsRNA intermediate [45,113]. The role of NS4A in flavivirus RNA replication was first reported in the Kunjin virus[34]. The topology of predominantly alpha-helical NS4A protein has three predicted transmembrane helices (56-126) and an N-terminal cytoplasmic helix with a kink which causes a bend in the helix (1 to 56 residues). The criticality of NS4A in virus replication comes from several mutational studies where most of the mutations reside on the N-terminal cytoplasmic helix affecting viral replication and protein stability; other than these, N-terminal alpha-helices remaining parts of NS4A are in ER membranes, and lumen side showed to interact with specific lipids with their conserved motifs [206].

Many host protein interactions mediate the membrane insertion of NS4A; NS4A interacts with signal recognition particle receptor and sec61G specifically, which is involved in the

co-translational insertion of transmembrane proteins explicitly with NS4A protein of DENV and ZIKV[207]. Recent studies showed TMEM41B, a lipid scramblase that resides on the ER membrane, will interact with both NS4B and NS4A, and another protein named ER membrane protein complex (EMC), a chaperon helping NS4A and NS4B for its correct topological arrangement on ER membrane otherwise which leads to degradation of these proteins[179,208]. Later NS4A has incorporated into ER membrane other host proteins, most notably the reticulon 3.1a protein, which does membrane rearrangement and bends the membrane to give outlier to replication vesicle and vimentin polymers (type of intermediate filament proteins) help to hold the vesicle's shape and anchor other proteins inside the replication vesicle by interacting with NS4A [209,210]. All these host proteins residing on ER membrane may be possible through direct interaction or passive diffusion in the ER membrane interacting with NS4A. Furthermore, NS4A cytosolic N-terminal 50 residues shared physical interaction with the soluble protein vimentin.

Among all nonstructural proteins, NS4A was the one that interacts with most of the host proteins along with these ER residing proteins and affects different pathways to promote viral survival, and inhibits innate immune responses of host cells against the virus. NS4A interacts with critical proteins in developmental stages like ANKEL2 in neuroblast cells, which is involved in nuclear envelope rearrangement during mitotic exit; this interaction affects downstream pathways of ANKEL2 during development stages and causes severe neuropathogenesis(microcephaly and eye size in ZIKV infection) [207,211]. NS4A also inhibits interferon-mediated innate immune response by competing with viral dsRNA detecting helicases like retinoic acid-inducible gene I (RIG-I) and DDX42 (with CARD domain). NS4A also interacts with their downstream adaptor protein in the pathway called mitochondrial antiviral-signaling (MAVS) protein (which regulates interferon-Ibased

innate immune response), and NS4A also inhibits the production of IFN-1 by acting on interferon promoter elements. Furthermore, NS4A interfering in the JAK-STAT pathway leads to developmental disorders, inhibition of innate immune responses, and induces autophagy[118,212]. Moreover, NS4A alone is enough to induce PI3K-dependent autophagy, which causes upregulation of pro-survival signalling, saves cells from death, and promotes viral replication inside the cell[56]. Furthermore, upregulation and manipulation of unfolded protein response proteins were reported for NS4A-expressed cells, and another host protein interaction of NS4A with PTB is essential for the first step in genome replication (for negative-sense RNA synthesis)[206].

Viral-coded other partner NS protein interactions with NS4A are essential for replication and viral multiplication inside the cell. A trans complement mutational study in yellow fever virus showed that mutations in NS1 protein affect the synthesis of negative-sense RNA had suppressor mutations in NS4A, indicating their genetic interaction[213]. Physical interaction has been observed between NS4B-NS3, NS4A-NS4B, NS3-NS5[143,147,205], and NS1 with NS4A, NS4B[160,214]. Furthermore, several other studies have shown NS2B, NS3, and NS4A enrichment[34] in replication vesicle and their colocalization[215]. And a recent study by Cortese. et al. showed the possibility of NS2B3 and NS4A-2K4B interaction for replication organelle synthesis[216].

Most of the NS protein processing from polyprotein is mediated by NS2B3 protease. A recent study showed that early events of this polyprotein processing happen near the rough endoplasmic reticulum (RER) surface, and this process is transferred into virus induced convoluted membranes to maintain the homeostasis of NS proteins inside the cell by ER stress-induced proteasome-mediated degradation; inhibition of this process

showed decreased viral genome replication [4,43]. After polyprotein processing, all the proteins needed for replication reach the ER membrane, where inward, spherical replication vesicles are present. On the polyprotein, a signal peptide: 2K positioned between NS4A and NS4B will help these two proteins reach the ER membrane. This region also helps NS4A induce membrane alteration in ER membrane in other flavivirus members, but in DENV, NS4A alone induces membrane alterations[113,217]. Other NS proteins that lack signal peptides may colocalize to ER membrane by interacting with NS4A or NS4B. Both NS4A and NS4B form homo-oligomers, and genetic interactions between NS4A and NS1 or NS4B were reported for yellow fever virus (YFV) and WNV, which are closely related flaviviruses[45,143,213,218].

NS4A cytosolic N-terminal region (1-50aa) has a role in modulation of NS3 ATPase activity was reported for WNV and ZIKA virus (ZIKV), and the acidic EELPD/E motif which presents towards the c-terminal of the cytoplasmic loop was shown to be responsible for this modulation this may be mediated by a transient intermediate of uncleaved NS3-NS4A [140,158]. This NS3-NS4A transient intermediate is needed for NS4BNS5 cleavage that gives functional NS5 protein indicating that NS3-4A may be the last

cleavage incident in the polyprotein processing[73,93,120]. Other than this interaction due to an uncleaved intermediate, the possible interaction between NS3 and NS4 may share a different interface and function differently after the cleavage event. NS3 may possibly interact with NS4A2KNS4B and colocalize to the replication organelle, where it will form a replication complex [216]. In NS3, motifs I, II, and VI are involved in ATP binding and hydrolysis, which are structurally positioned towards the protease domain. In the NS2B bound form of NS3, if NS4A interacts with one of these ATP binding motifs with its acidic

motif remaining cytosolic NS4A N-terminal may have the chance to interact with the protease domain, which is proximally positioned to these ATP binding motifs, and it may interact with either the protease domain or the NS2B cofactor region.

No drugs are available that target NS4A, a vital protein with many functions; characterization of NS4A and establishing its dynamic interaction with other nonstructural proteins will provide more drug targets. With these clues, we started our study to characterize these two proteins, NS4A and NS2B3 protease proteins, establish their physical interaction, and functional characterization of this interaction. T

This study first characterized the novel interaction between NS4A protein and NS2B3 protease through different molecular biology methods. Bacterial two-hybrid assay results showed that the primary determinant for interaction with NS2B3 protease is within the N-terminal 56 residues region of NS4A. Co-expression, far-UV CD spectroscopy, size exclusion chromatography, and molecular docking studies, using predicted structures of NS2B3 and NS4A, corroborated the bacterial two-hybrid assays and explained the nature of interactions.

Moreover, established interaction between NS2B3 and NS4AN56 and observed that the interaction of NS4AN56 negatively affected the protease activity of NS2B3 protease. So possibly NS2B3, after interaction with NS4A, should switch its function from polyprotein processing to replication for genomic RNA secondary structures unwinding during replication.

3.2 Methods and Materials

3.2.1 Dengue NS4A Sequence Analysis and Ab-Initio Structure Prediction of NS4A and Molecular Docking

The sequences of NS proteins of the Dengue virus used in this study are from NCBI GenBank accession number JN903579. Transmembrane helices predictions on the sequences are made with TMHMM, Phobius, or PSIPRED MEMSAT-SVM by submitting the NS4A sequence on their respective servers.

NS4A ab-initio model was generated using alphafold2[173] colab notebook (https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb, shared publicly through a Creative Commons Attribution-NonCommercial 4.0 International license) and RoseTTAfold[174] (https://robetta.bakerlab.org/myqueue.php). This AlphaFold Colab notebook uses the BFD database for prediction after a multiple-sequence alignment. After an energy minimization step, the predicted model PDB file is downloaded along with the pLDDT scores. Structure Figures were prepared with PyMOL. Since there is no significant difference in the predictions using AlphaFold2.0 or RoseTTAfold, we used the structure predicted by AlphaFold2.0 for further analysis.

For molecular docking simulations, we used the energy-minimized structure of the homology model that we built for DENV serotype-1 NS2B3 protease of both open and closed conformations, and NS4A N-terminal 56 residues region of AlphaFold predicted structure. Molecular docking simulation was performed on a rigid body; global, blind docking was run on ClusPro server[176] (https://cluspro.bu.edu) to get a possible docking pose and information about the interacting residues.

3.2.2 Bacterial Two-Hybrid Assay

A bacterial two-hybrid assay (BACTH) was performed following the protocol described earlier[219]. Full NS4A or NS4AN56 only (residues 1-56) are cloned as a C-terminal fusion with the T25 fragment of the adenylate cyclase in the BACTH assay system pKT25 plasmid. NS2B cofactor region of (residues 48-100) region, NS3 protease (residues 1178) or a chimera of these region linked by glycine serine linker (GGGSGGG) named as NS2B3[151], from DENV serotype-1 is cloned as N-terminal fusion of the T18 fragment of the adenylate cyclase in pUT18 vector. To perform the BACTH assay, different combinations of pKT25-NS4A helicase and pUT18-NS2B3 plasmids are used for transforming the E.coli BTH101 strain. The transformant cells were plated on an LB agar plate containing 50 µg/ml kanamycin and 100 µg/ml ampicillin. The transformants that grew on the plate were patch streaked on MacConkey-maltose Agar indicator plates supplemented with 50 µg/ml kanamycin, 100 µg/ml ampicillin, and 0.5 mM IPTG and incubated at 30^{o} C for 96 hours. The plasmids pUT18-Zip, and pKT25-Zip, containing interacting domains of the GCN4 leucine zipper, were used as positive control in the assay. Empty pUT18 and pKT25 plasmids were used as a negative control for the assay.

3.2.4 Cloning of NS2B3, NS3CTD-NS4AN56, GFP-Cs-Quencher, NS4A, NS4AN56 Into pET24b Vector and pETduet Vector

NS4A of Dengue serotype-1 (DENV1) with sequence Gene Bank accession number of JN903579 was previously cloned into pGEM-T vectors and named pDV1-419NS4A. DNA coding region of NS4A full length and N-terminal 56 residues region of NS4A alone were PCR amplified (primers SM7 and SM8) and cloned in between NdeI and XhoI sites of pET24b vector with a hexahistidine tag. For co-expression studies, a pET Dute construct of NS2B3 and NS4AN56 was made by subcloning NS2B3 into the first MCS

in between BamHI and HindIII with an N-terminal hexahistidine tag. Furthermore, NS4AN56 into 2nd MCS of pET-Duet, between NdeI and XhoI with an extension of an S-Tag (Lys-Glu-Thr-Ala-Ala-Ala-Lys-Phe-Glu-Arg-Gln-His-Met-Asp-Ser) coding region towards C-terminal of the NS4AN56, and it was labeled as pET Duet NS2B3NS4AN56 vector.

The cloning methodology of NS2B3, NS3CTD-NS4AN56, and GFP-Cs-Quencher method was explained in the second chapter. (Detailed in method and material section 2.2.1).

3.2.3 Bacterial Expression and Purification of NS2B3, NS3CTD-NS4AN56, GFPCs-Quencher, NS4A, NS4AN56 Recombinant Proteins

pET24b vector containing NS4A gene was transformed into C41DE3 cells, and cells were grown in LB Broth supplemented with 50µg ml^{-1} kanamycin to an optical density at 600 nm(OD_{600}) of 0.6 to 1.0 at 37 ºC. Expression of recombinant NS4A was induced with 1mM isopropyl β-D-1thiogalactopyranoside (IPTG) and incubated at 37ºC for 4 hours. Cells were pelleted by centrifugation at 5000 × g for 10 mins at 4ºC. Pellets were resuspended in a lysis buffer of 20mM Tris pH 7.4 containing 150mM NaCl and disrupted by sonication using a digital probe sonicator at 40% amplitude for 20 min with 15 seconds off and 5 seconds on cycles. The cell lysate was clarified by centrifuging at 20000 × g at 4º C for 20 min, and the protein was expressed in the pellet fraction in the membrane bound form. The pellet fraction was taken for further purification by solubilizing it in a detergent with a buffer of 20mM Tris pH7.4 containing 3% LDAO (lauryl dimethyl amine oxide) 150mM NaCl and 20mM Imidazole overnight. Solubilized protein was clarified by centrifugation at 20000 × g at 4º C for 20 min and resulted in the supernatant was diluted to 0.5% LDAO with 20mM Tris pH 7.4 containing 150mM NaCl and then loaded

on Ni-NTA gravity column which was pre-equilibrated with 0.5% LDAO with 20mM Tris pH 7.4 containing 150mM NaCl buffer. After the sample was loaded onto the column, the column was washed with 5 column volumes of wash buffer of 20mM Tris pH7.4 containing 0.5% LDAO, 150M NaCl, and 50mM Imidazole. Then bound protein was eluted with elution buffer 20mM Tris pH 7.4 containing 0.5% LDAO, 150mM NaCl, 300mM Imidazole. Eluted protein was concentrated and loaded onto the superdex-75 column for further purification, a protein purified after this step was stored at -20^{o} C till further use. pET24b vector containing the NS4AN56 gene was transformed into BL21DE3 cells, and

cells were grown in LB Broth supplemented with 50µg ml^{-1} kanamycin and 34µg ml^{-1} chloramphenicol to an optical density at 600 nm (OD_{600}) of 0.6 to 1.0 at 37^{o} C. Expression of recombinant NS4AN56 was induced with 1mM isopropyl β-D-1thiogalactopyranoside (IPTG) and incubated at 37^{o} C for 4 hours. Cells were pelleted by centrifugation at 5000 × g for 10 mins at 4ºC. Pellets were resuspended in a lysis buffer of 20mM Tris pH 7.4 containing 150mM NaCl (TN Buffer) and disrupted by sonication using a digital probe sonicator at 40% amplitude for 20 min with 15 seconds off and 5 seconds on cycles. The cell lysate was clarified by centrifuging at 20000 × g at 4^{o} C for 20 min, and the protein expressed is in a soluble fraction. Clarified cell lysate containing expressed protein was loaded on the Ni-NTA gravity column, pre-equilibrated with TN buffer. After the sample was loaded onto the column, the column was washed with 5 column volumes of wash buffer of 20mM Tris pH7.4 containing 150mM NaCl and 50mM Imidazole. Then bound protein was eluted with elution buffer 20mM Tris pH 7.4 containing 150mM NaCl and

300mM Imidazole. Eluted protein was concentrated and loaded onto the superdex-75 column for further purification, a protein purified after this step was stored at -20 ºC till further use.

The methodology of expression and purification of NS2B3, NS3CTD-NS4AN56, and GFP-Cs-Quencher proteins method was explained in the second chapter (detailed in method and material section 2.2.2).

3.2.4 Co-expression of NS2B3-NS4AN56 for Interaction

To test the Interaction between NS4AN56 and NS2B3 protease, we cloned both genes into the pET Duet vector. The NS2B3 protease coding DNA cloned into MCS having the N-terminal His tag and NS4AN56 coding DNA cloned into 2nd MCS with S-tag, the resulted proteins expressed from this construct will have NS4AN56 with C-terminal Stag, and NS2B3 has an N-terminal hexahistidine tag.

If these two proteins interact when co-expressed, they will elute together in Hexahistidine tag affinity chromatography purification. Proteins were coexpressed and co-purified using His affinity chromatography method from Rosetta DE3 cells as it was described for NS2B3 alone.

3.2.5 Protease Assay

The methodology for protease assay was explained in Chapter 2 (detailed in section 2.2.3)

3.2.6 Size-Exclusion Chromatography

We used size exclusion chromatography (SEC) to analyze recombinant NS4A and NS4AN56 protein quaternary structure analysis. For NS4A protein SEC, we used superdex 200 Increase GL 10/300 column with 20mM Tris pH 7.4 containing 150mM NaCl and 0.05% of LDAO as running buffer. NS4AN56, NS2B3, and NS2B3

proteaseNS4AN56 SEC done using superdex-75 increase GL 10/300 column with 20mM Tris pH7.4 containing 150mM NaCl and 2mM EDTA as a buffer.

3.2.7 Protein Crystallization

NS4A protein extracted and purified using LDAO detergent was in 20mM Tris pH 7.5 containing 150mM NaCl and 0.01%LDAO final concentration buffer. Protein was concentrated at 10-15mg/ml concentration. 35% DMPC: CHAPSO mixture at a 2.8:1 molar ratio bicelles were prepared[220]. By varying bicelle concentration from 4-8% and protein concentration 5-12mg/ml, crystallization drops were set up in the hanging drop method with JCSG, MemFac, MemPlus, and Memgold conditions.

3.2.7 Far-UV Circular Dichroism (CD) and Intrinsic Tryptophan Fluorescence Spectroscopy

Far-UV CD spectra were recorded with NS4A, NS4A N56, NS2B3, and NS2B3NS4AN56 co-expressed purified complex proteins were done using Jasco 1500 spectropolarimeter. Samples were incubated at room temperature for 30 min before recording the spectra. Protein was used at 0.2 mg/ml concentration, and far-UV CD spectra were recorded using a 0.1 cm path length quartz cuvette from 200nm to 250nm wavelength in 20mM Tris pH 7.5 containing 150mM NaCl buffer (TN buffer). Normalized Mean Residue Ellipticity (MRE) values against wavelength were calculated, and spectra were plotted. For NS4A and NS4AN56 proteins, spectra were recorded in 20mM Tris pH 7.4 containing 150mM NaCl and 0.05% of LDAO buffer and 20mM Tris pH 7.4 containing 150mM NaCl buffer, respectively. For NS2B3 protease and NS2B3 protease-NS4AN56 co-expressed protein, we used two different buffers; 50mM Tris pH 9.0 containing 10% glycerol buffer (in which protease assays are done) and TN buffer.

The data was analyzed by converting the ellipticity(Θ) values to $Θ_{m.r.e}$. If there are any secondary structure changes, upon interaction, it is expected that the observed normalized $Θm_{.r.e}$ for the mixture would be different from the expected value.

Intrinsic tryptophan fluorescence spectroscopy to test the Calcium-binding of NS4A and NS2B3 protease proteins. Both the proteins were titrated with Calcium Chloride from 016mM in no salt 20mM Tris pH 7.5 buffer, and emission spectra were recorded using a 1 cm pathlength cuvette. Spectra were recorded using an FLS 1000 fluorimeter (Edinburgh Instruments, UK). The protein was used at 0.05 mg/ml concentration for recording the fluorescence spectra. The excitation wavelength was set to 295 nm, and the emission wavelength range was set to 300-500 nm. The excitation and emission bandwidths are set to 1 nm. Spectra were blank corrected and smoothened using the Fluoracle (Edinburgh Instruments, UK) software. For spectra plotting, respective $CaCl_2$ concentrations having buffer blanks without proteins were subtracted from the protein spectra with the same concentration of $CaCl_2$.

Parameters for CD spectra and fluorescence spectra collection and plotting was similar as it was explained in chapter one methods and materials (detailed in section 1.2.8).

3.3 Results

3.3.1 N-terminal Cytoplasmic Helix of NS4A Showed Interaction with NS2B3 in BACTH Assay

To determine and characterize the interaction between NS4A and NS2B3 protease, a bacterial adenylate cyclase-based two-hybrid assay (BACTH) was performed as a starting point. Before starting this assay, we did an initial sequence analysis and ab-initio structural characterization of NS4A to understand the protein and make different

truncations for its interaction studies. Previous studies established that NS4A is a transmembrane protein[113]. Two mark the boundaries of the transmembrane helices; we did sequence analysis in different secondary structures and transmembrane prediction servers. NS4A showed it is predominantly alpha-helical and has three transmembrane helices and a cytoplasmic helix. The cytoplasmic helix boundary for DENV-1 NS4A ranged between residues 1 to 56 in the TM prediction algorithms (Figure 3-1).

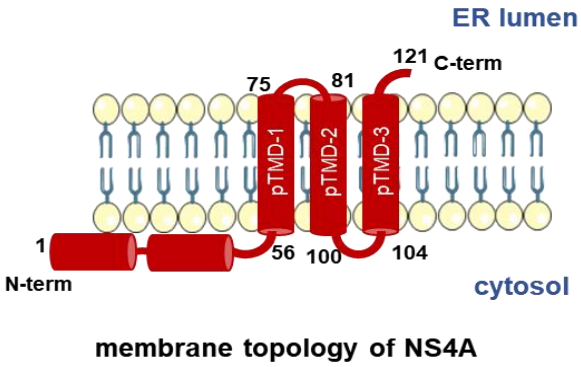


Figure 3-1 NS4A membrane topology model made based on predicted transmembrane helices from different prediction algorithms (TMHMM, Phobius, PSIPRED, and TM-PRED)

To get more information about the structural arrangement of these transmembrane and cytoplasmic cores of the protein, the NS4A protein sequence of DENV1 submitted to the recently introduced near-accurate structure prediction AI-based algorithms AlphaFold2[173] (through google colab book) and RoseTTAfold[174] server (robetta online server) and generated ab-initio structure models. DENV1 NS4A structure model generated by Alphafold2 was assessed by its pLTDD score, which is more than 80. Overall sequence-based structure analysis of the pLTDD score of NS4A showed loop regions between helices have lower scores than helices (more than 90). The RoseTTAfold model of DENV1 NS4A has a confidence score of 0.75, and the angstrom error rate for the entire sequence fluctuated between 2 and 3 angstroms, which is

significantly reliable. Both models overlapped when structures were aligned (except N-terminal 56 residues region).

The N-terminal helix (ranging from 1-56 aa region) is away from the transmembrane core and is not predicted as a transmembrane helix in any transmembrane prediction algorithms indicating that this helix may be posed towards the cytosolic side of the ER membrane in the replication complex. To further validate this DENV1 NS4A fold or predicted structure, we submitted an NS4A sequence of different flaviviruses like WNV, ZIKV, and YFV (which have a similar sequence) to the RoseTTAfold and compared their predicted structure, forming an identical fold or not (Figure 3-2). NS4A Ab-initio models generated for these viruses were compared by structural alignment using PyMol. All predicted ab-initio structures have an identical transmembrane core with flexible Nterminal helix DENV-1 NS4A.

Based on this sequence and ab-initio generated structural understanding, two different constructs of NS4A were made for BACTH assay: NS4A full length and N-terminal 56 residues alone; this N-terminal truncation named NS4AN56 (In TM-Pred server, the first transmembrane helix, started from 56th residue).

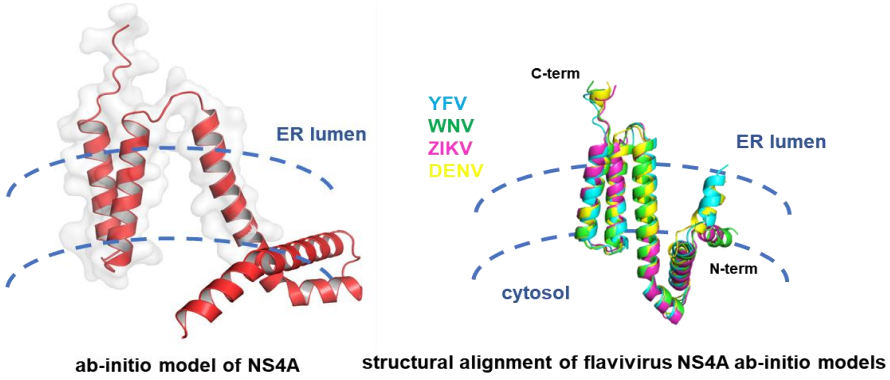


Figure 3-2 Alphafold2-based Ab-initio predicted structure of NS4A (red) and Structural alignment of NS4A structure predicted for different flaviviruses

For interaction identification and to map the interaction, different constructs of the NS2B3 protease were made: NS2B cofactor (48-100), NS3 protease (1-173), and NS2B3 protease construct having cofactor region of NS2B (48-100) and NS3 protease (1-173) domain linked with a glycines linker[151] (similar construct reported previously used for several other studies), for BACTH assay.

The BATCH assay was conducted by double-transforming these plasmids in two different backgrounds, and cells were plated on MacConkey/maltose agar indicator plates. In the BACTH assay, the interaction between proteins was assessed qualitatively by red color development in double-transformed bacterial colonies on MacConkey/maltose agar indicator plates compared to color development in control plates. The robust color development is observed with N-terminal 56 residues of NS4A compared to full-length NS4A (Figure 3-3) with NS2B cofactor, NS3 protease, and NS2B3 protease.

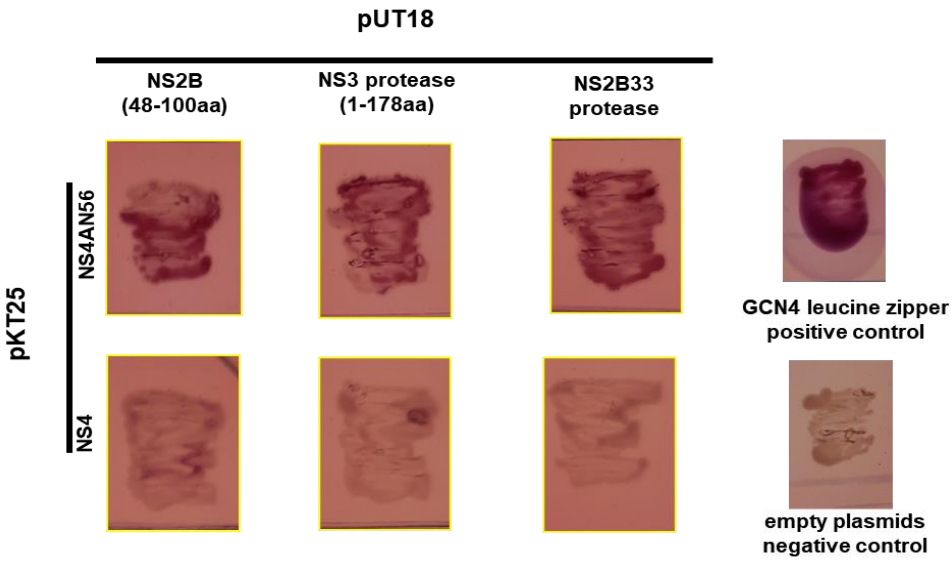


Figure 3-3 BACTH assay of different truncations of NS4A against NS2B3 protease different domains

3.3.2 NS4A is Predominantly an Alpha-Helical Protein That Forms a Homomeric Trimer, and Both NS2B3

To further validate our BACTH results, DENV serotype 1 NS4A, NS4AN56, and NS2B3 protease proteins were expressed and purified in a bacterial expression system, and a thorough biochemical characterization of these purified NS4A and NS2B3 protease proteins was done before going for interaction studies.

NS4A full-length, NS4AN56 and NS2B3 protease coding DNA constructs were cloned into a bacterial expression vector pET24b with C-terminal his tag. Initial NS4A fulllength protein expression optimization trials were done in different strains and got better expression in C41DE3 cells (for remaining strains like BL21DE3 protein was toxic and expression significantly less), and the NS4A protein expressed is in the cell membrane fraction. Protein extraction from membrane fraction was done using detergent LDAO. Protein was purified in the presence of LDAO in a two-step purification method, including hexahistidine tag affinity chromatography and gel-filtration chromatography in 0.05% LDAO (Figure 3-4A). After Ni-NTA chromatography purification, eluted protein in ~90% pure, in gel-filtration chromatography protein eluted as a single sharp peak at elution volume corresponding to a globular mass of ~66kDa, forming a homomeric oligomer (monomeric mass of protein ~15kDa) (Figure3-4C) and this homomeric oligomer form of purified NS4A was reported previously[115]. After purification, protein buffer was exchanged into a 0.01% LDAO containing TN buffer and stored at a 5 mg per ml concentration in the freezer. When we tried to avoid detergent, protein got aggregated, conforming to its absolute detergent requirement for its stability. Secondary structure analysis by far-UV circular dichroism spectroscopy of NS4A shows a spectral signature

(Figure 3-4B) typical of a predominantly helical protein with negative peaks at 208 and 222nm, consistent with our sequence analysis and ab-initio structure model.

Unlike NS4A full-length protein, NS4AN56 protein was expressed in BL21DE3 cells in the soluble fraction. The expressed protein was purified in native conditions in a two-step purification method, including a hexa-histidine tag affinity chromatography followed by gel-filtration chromatography (Figure 3-4A). The purified protein is eluted in gel-filtration chromatography as a single sharp peak at an elution volume corresponding to a globular mass of ~21kDa, possibly forming a homomeric trimer (Figure 3-4D). Secondary structure analysis by far-UV circular dichroism spectroscopy shows a spectral signature (Figure 3-4B) typical of a predominantly helical protein with signature negative peaks at 208 and 222nm. With this predominant soluble alpha-helical nature of NS4AN56, there is no need for any cosolvent and detergent, confirming its cytosolic positioning.

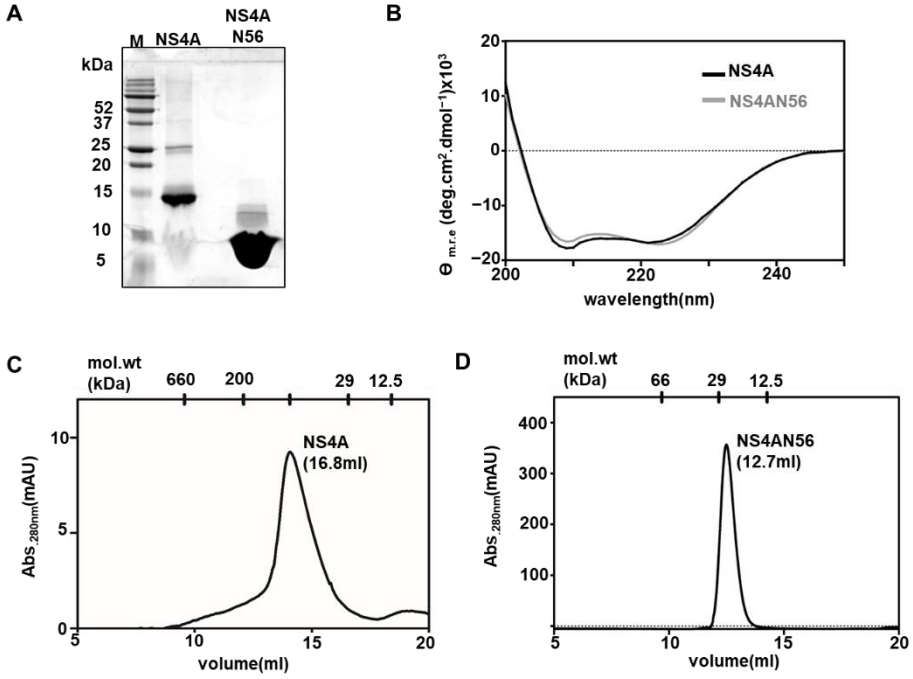


Figure 3-4 Recombinant NS4A, NS4AN56 proteins purification and secondary structure analysis (A) SDS-PAGE of purified recombinant NS4A and NS4AN56 (B) Far-UV-CD Spectra of NS4A and NS4A N56 (C) Size exclusion chromatogram of NS4A (D) Size exclusion chromatogram of NS4AN56

Crystallization of NS4A full-length protein was attempted in the presence of detergent and bicelles, but we got only fragile needle-shaped crystals. Further optimization with seeding and detergent did not give any diffraction-quality crystals.

NS2B3 protease and its substrate expression and purification and the protease assay results were described previously in Chapter 2.

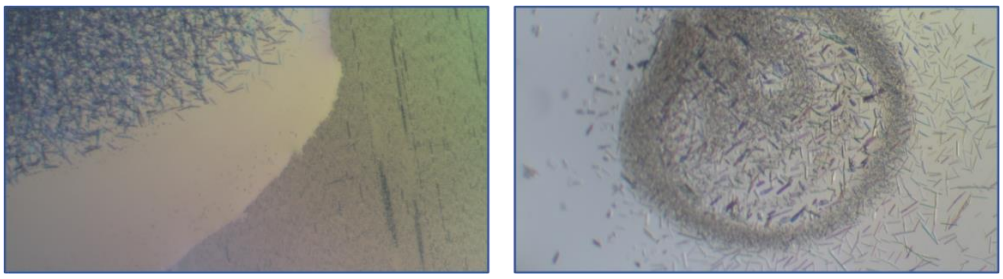


Figure 3-5 NS4A crystallization, Needle-shaped NS4A crystals in bicells

3.3.3 NS4AN56 Interacted With NS2B3 Protease When Co-expressed Together

NS4AN56 protein was used for interaction studies; the rationale for proceeding with NS4AN56 is that the remaining part of the protein will be in the membrane and inaccessible for any soluble protein interaction. Moreover, our BACTH assay showed that the N-terminal 56 residues are enough for interaction. To further validate our BACTH assay results, we took the purified proteins NS4AN56 (trimer) and NS2B3 protease, mixed them in a 1:1 molar ratio, and loaded them onto the SEC column to see co-elution or complex formation. However, we did not see any significant shift from corresponding control peaks which loaded independently the same amount as in (Figure 3-6B).

It was reported that an uncleaved NS3-NS4A transient intermediate during polyprotein processing was required for NS4B-NS5 cleavage[206], and results from the BACTH assay gave indications that this interaction may seem co-translationally or during polyprotein processing. Moreover, the independently expressed NS4AN56 protein homo-oligomers interface may hide these two proteins' interaction interface. Moreover, NS3 protease and NS2B interaction studies were established with glycine-linked protein and dual expression system. In a similar way, NS4A and NS2B3 protease were cloned into two different pET duet vector MCS sites as two different ORFs, one with hexahistidine tag (N-terminal of NS2B3 protease), the other without tag (although an S-Tag present towards C-terminal of NS4AN56). In this co-expression study, if these two proteins interact, they will co-purify from the Ni-NTA His-tag affinity column. In the SDS-PAGE analysis, an extra band other than NS2B3 protease indicates that both proteins got copurified and formed a complex (Figure 3-6A).

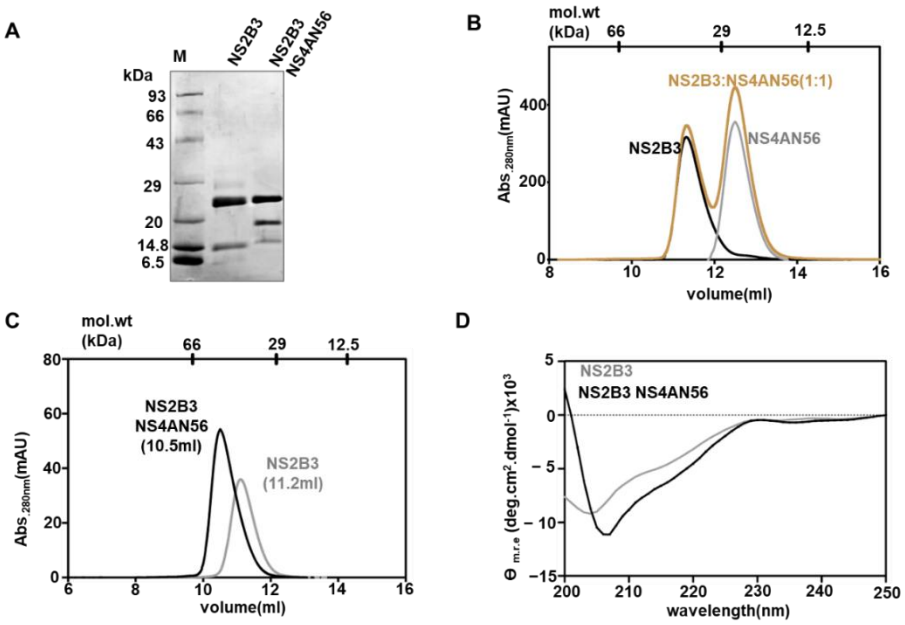


Figure 3-6 NS2B-NS4AN56 interaction (A)SDS-PAGE of recombinant NS2B3 and NS2B3NS4AN56 (Co-expressed) proteins (B)SEC of NS2B3 (black), NS4AN56(grey), and

NS2B3:NS4AN56 in 1:1 molar ratio (gold) (C) SEC of NS2B3 and NS2B3-NS4AN56 (co-expressed) (D) Far UV CD spectroscopy of NS2B3 and NS2B3-NS4AN56 proteins

SEC is performed with NS2B3 and NS4AN56 as controls to reconfirm this interaction and complex formation. In SEC NS2B3-NS4AN56 eluted as a single sharp peak, the retention volume of this co-purified protein showed a significant shift to control of NS2B3 protease (Figure 3-6C). Furthermore, in far-UV CD spectra comparison of NS2B3 protease, NS2B3-NS4N56 has a significantly different signature with a negative peak at 208nm; NS2B3 protease has a typical chymotrypsin fold having protein spectra and NS2B3-NS4AN56 protein with the shift of 205nm negative peak to 209 nm (Figure 36D) given NS4AN56 contribution in the overall spectra change.

3.3.4 NS4AN56 Interaction Showed Reduced Protease Activity of NS2B3 Protease

Protease assay of NS2B3 protease showed no further increase in the activity after 2 hours. Functional characterization of this protein interaction, and to test the effect of NS4AN56 on NS2B3 protease activity, we did a protease assay with NS3CTD-NS4AN56 substrate. In the assay, the intensity of two cleavage products at their respective molecular weights ~17kDa and ~7kDa after a specific cleavage event was qualitatively assessed to see the activity. NS2B3-NS4AN56 protease activity was very low, with no second cleavage product trace (~7kDa) (Figure 3-7B).

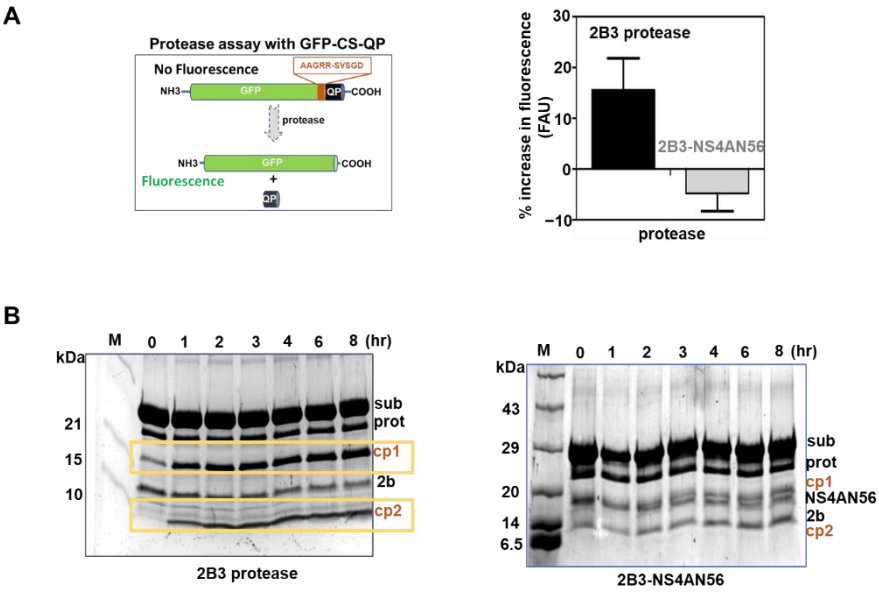


Figure 3-7 Protease assay of NS2B3-NS4AN56 (A) Schematic of protease assay with GFP-CS-QP substrate comparative graph percent cleavage with NS2B3 protease and NS2B3-NS4AN56 protease (B) SDS-PAGE of protease assay with NS3CTD-NS4AN56 substrate

To further confirm this negative regulation or inhibition of NS2B3 protease, a protease assay with a fluorescence reporter-based substrate was designed where a GFP linked to quencher peptide with protease cleavable linker representing junction between NS3NS4A. In the assay, after protease-mediated cleavage, the quenching peptide will be away from GFP, leading to the increased emission peak intensity of GFP, which will be taken as a measurement of protease activity. A similar assay was used previously[221]. Assay conditions were the same as previously mentioned, with an endpoint after 9 hours, in this assay also saw very low or no activity of NS2B3 protease in the presence of NS4AN56 (Figure 3-7A).

To map the interaction and to deduce how this interaction inhibits the protease activity a molecular docking simulation in the cluspro web server (a global docking server) was performed. In this study, NS2B3 protease open and closed conformations homology

models are taken as receptors and NS4AN56 as a ligand (Ab-initio structure generated from AlphaFold2) to map this interaction interface between these two proteins and deduce the possible reason behind the loss of activity. A docking pose with a significant cluster size with the lowest energy score with both open and closed conformations showed the interaction of NS4AN56 near the catalytic pocket of NS2B3 protease and masking this catalytic pocket partially (Figure 3-8).

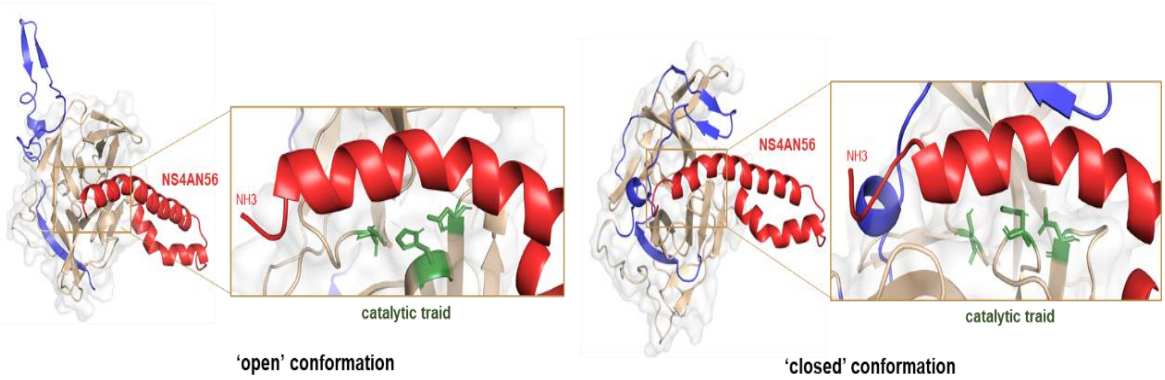


Figure 3-8 Molecular docking of NS4AN56 with NS2B3 protease (red colour is NS4A, wheat colour cartoon is NS3 protease domain, blue colour is 2B co-factor region, green sticky representations are catalytic traid)

3.4 Discussion

In RNA viruses, the switch from Polyprotein processing to replication is a temporal event, along with sorting replication complex into replication vesicles. Sorting of the replication complex into ER membrane might need a signal peptide. The only known signal peptide '2K' is between NS4A and NS4B. So, it is the only possible way to guide the remaining nonstructural; proteins to ER through interacting with these core proteins. Interaction between NS4B-NS3, NS3-NS2B, NS4A-NS4B, NS1 with NS4A2KNS4B, and NS3-NS5 have been established previously through several studies. Mapping the interaction interfaces and defining the functional importance of these interactions will open doors for structure and function-based drug designing approaches.

Based on the initial sequence analysis and ab-initio generated structure models, different domain truncations were made in NS4A and NS2B3 protease for bacterial adenylate cyclase two-hybrid assay to test the interaction. In the BACTH assay, the interaction of NS4AN56 showed robust color development with NS2B, NS3 protease, and NS2B3 protease, given indications that interaction is spread between NS2B and NS3 protease. A recombinant protein-mediated invitro approach was taken to validate this and map the interaction interface further.

This present study conducted a thorough biophysical and biochemical characterization of NS4A. The initial sequence and ab-initio structure model generated by Alphafold2 and Rosettafold have identical folds and match the topology predicted previously by an NMR study[222] with a transmembrane core of three membrane-spanning helices (56-120) and a flexible N-terminal helix region away from this transmembrane core. This is further confirmed by our recombinant expression and purification, where both NS44A and NS4AN56 proteins require different expression and purification approaches. Purified NS4A protein forms a homomeric oligomer in the presence of detergent. Similarly, soluble NS4AN56 also formed a trimer, this is conformed NS4A homomeric oligomer mediated by N-terminal 56 residues region, and it is a trimer. Moreover, compiling the data from both our ab-initio model and biophysical studies confirms its topological arrangement on ER membrane that N-terminal 56 residues are towards the cytosolic side of the ER membrane remaining part is in the membrane.

After this core, structural and functional understanding and initial clues from the biophysical characterization of both the proteins and bacterial two-hybrid assays gave initial information regarding the Interaction between NS2B3 protease and the NS4AN56

region. And the NS4AN56 interaction interface may spread between the NS2B cofactor and an NS3 protease domain interaction interface. To confirm physical interaction and fine map the interaction interface between these two proteins.

NS4AN56 alone for interaction studies because our initial sequence and structural analysis confirmed later part from residues 56-120 is inside the membrane; adding to that, the full-length NS4A protein requirement of detergent for its stability may interfere with our interaction and biochemical studies. Initially, trials test the interaction by expressing and purifying the proteins separately and mixing them to see the complex formation. However, surprisingly, there was no complex formation; however, NS3 and NS2B interaction was established by linked proteins and co-expressing in a dual expression system. So, we took a similar approach, NS4AN56 showed interaction with NS2B3 protease when they co-expressed together in a bacterial dual expression system, and the SDS-PAGE analysis of purified protein from these cells showed three distinct bands conforming to complex formation. This is the first evidence of physical interaction between NS2B3 protease and NS4A protein. NS2B3-NS4AN56 complex formation was further confirmed by SEC, and secondary structural content change in far-UV CD spectroscopy compared to NS2b3 protease alone and complex formation between NS2B3 protease and NS4AN56. When we tested the effect of NS4AN56 on NS2B3 activity, to our surprise, the NS4AN56 interaction reduced the NS2B3 protease activity with both NS3CTD-NS4AN56 GFP-Cs-Quencher substrates. Later our molecular docking simulation studies showed NS4AN56 interacted near the catalytic pocket and masked it in closed and open conformations.

In other viruses belonging to the Flaviviridae family other than the Flavivirus genus, NS4A's positional arrangement on polyprotein is the same, but it is a different sequence and smaller than Flavivirus NS4A and has functions identical to NS2B cofactor flavivirus genus. Dynamic interactions between viral-coded proteases were reported for several RNA viruses. In norovirus, protease activity of polyprotein processing is regulated temporally in replication complex[46]. In Pestivirus, which belongs same Flaviviridae family, the switch from genome replication to particle synthesis happens by alternative protein-protein interaction between the NS4A cofactor and protease domain of NS3 [223]. Having multiple roles which are parted away in different structural domains gave NS3's extended role in flavivirus replication with different protein interactions.

The requirement NS3-NS4A transient intermediate for NS4B-NS5 cleavage gives clues that NS3-4A may be the last cleavage incident in the polyprotein processing[73,93,120] and several host protein interactions of NS4a which helps its correct localization topological arrangement in replication vesicles on ER membrane indicating NS4A may take a posttranslational path after its cleavage from polyprotein to ER membrane. After this cleavage event, NS3 may interact with NS4A2KNS4B and colocalize to the replication organelle, where it will form a replication complex [216]. After polyprotein processing, there is an ER-associated degradation mechanism in the viral-induced convoluted membranes that maintained the optimum levels (homeostasis required for viral replication) of nonstructural inside the cell; through this mechanism, NS4a was maintained at a very low level in the cell, inhibition of this process affected the viral replication[43]. This indicates a requirement for controlling the expression levels of processed nonstructural proteins inside the cells, most importantly NS4A, which negatively affects the viral replication in an unknown mechanism.

Although it is believed that polyprotein processing will happen near ER surface, there is no evidence for the exact spatial marking of polyprotein processing in Flavivirus. Previous literature confirmed that NS4A is an absolute requirement for viral replication complex formation based on its ER residing and cytosolic host protein interaction. And higher levels of NS4A inside the cells have a negative impact on viral replication.

Interaction and functional studies of NS2B3-NS4AN56 interaction and interaction interface mapping near the catalytic pocket of NS2B3 protease, the docking studies lead us to believe this NS4AN56 interaction induced negative regulation of NS2B3 protease activity may be a 'switch from polyprotein processing to replication' (Figure 3-9).

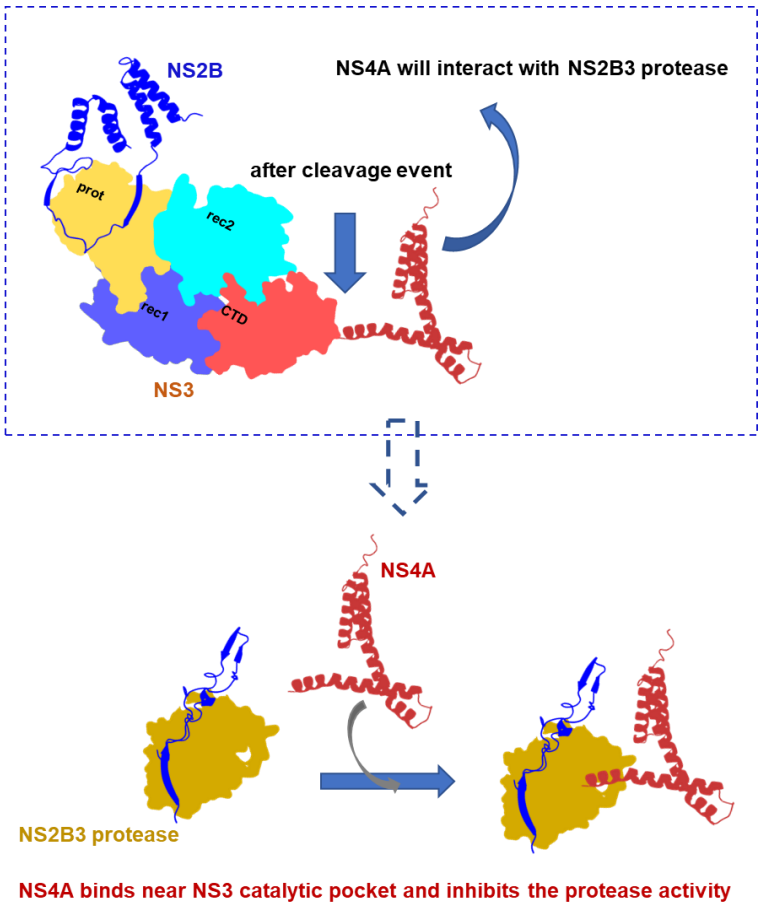


Figure 3-9 Schematic model for how NS4A interacts with NS2B3

Moreover, the interaction between NS3-NS4A is preliminary. It requires further validation and functional characterization in-vivo.

Conclusions

Concluding Remarks

The receptor preference for entry into the cell for disease-causing viruses belonging to the flavivirus genus differs. However, the molecular mechanisms used for replication and polyprotein processing have notable similarities. Understanding one of the representative viruses from the flavivirus genus, their replication, and polyprotein protein processing events can be extrapolated to the other viruses. Specifically, for DENV, due to different serotypes and immunopathology, developing a vaccine against it is a difficult task. After viral genome entry into the host cytosol, RNA genome translation into a polyprotein, later this polyprotein processing, and then replication are the next crucial steps for viral propagation and multiplication inside the host cell. Polyprotein processing will give the functional proteins and may be responsible for some of the interaction between the nonstructural proteins, which will initiate replication complex formation and regulate some crucial steps spatially and temporally [224].

Both polyprotein processing and replication are critical steps for viral multiplication. These two molecular processes involve all viral-coded nonstructural proteins and host proteins. NS3 is the one protein with multiple functions involved in these two processes, and with most nonstructural protein interactions, it is central to the replication complex. This study focused on such nonstructural protein interactions of NS3, mainly NS-NS4B, NS3-NS2B, and NS3-NS4A.

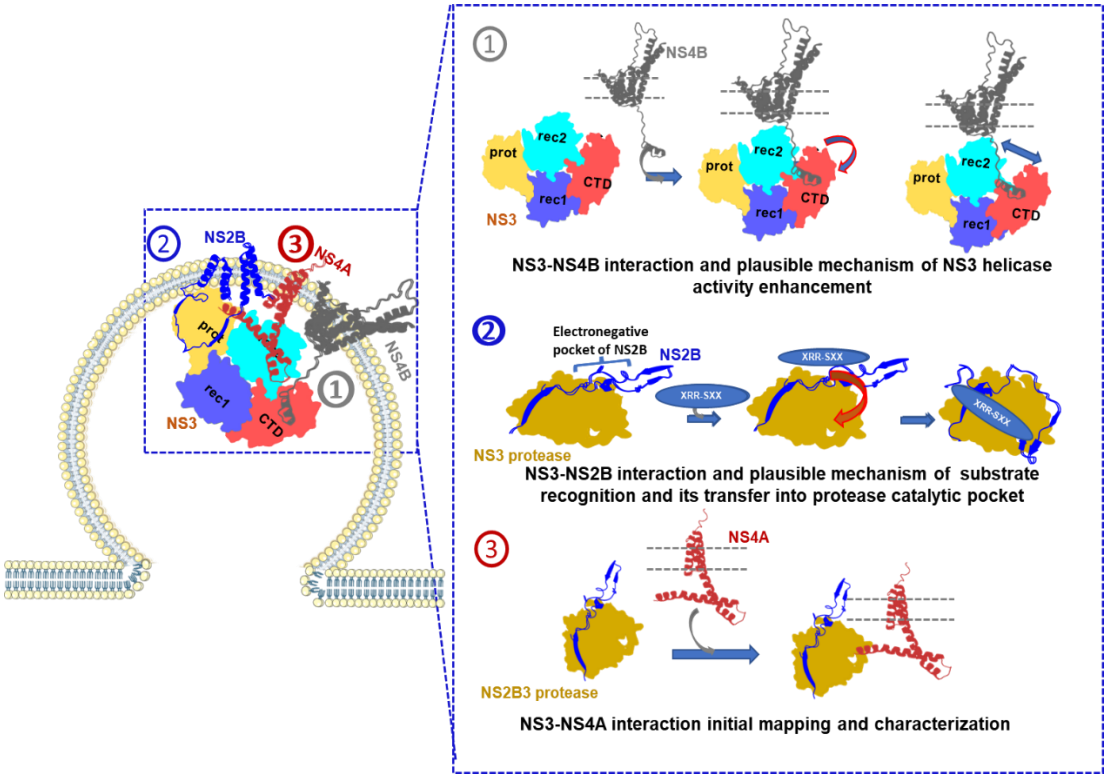
(1) In chapter one, NS3-NS4B interaction was mapped, and from all results taken together; a plausible explanation is that NS4B interaction with NS3 helicase increases the RNA duplex unwinding activity by increasing the CTD motions (towards and away from the RecA domains for each cycle of unwinding/RNA translocation). Furthermore, proposed a mechanism for the concerted action of

NS3-interacting protein complexes in viral negative-strand synthesis. The N-terminal disordered region of NS4B translocates into the cytosolic side of the RO after the proteolytic processing of the 2K C-terminus signal. NS3 interacts with the NS4B N-terminus region, which tethers it to the membrane. The NS4B Nterminus region wraps around the recA2 and CTD subdomains of the helicase and modulates the subdomain motions during the dsRNA unwinding step. NS3 also interacts with NS5 through the CTD. Through this interaction, the NS5 is also positioned on the 3′-end of the genome, where negative-strand synthesis starts from. As a result of these interactions, a tripartite complex of NS3–NS4B–NS5 may form in the RC, where the NS3–NS4B complex will precede NS5. The conserved stem-loop of the 3′-SL (in 3′-UTR domain III) is then unwound by NS3 helicase. NS4B modulates the helicase activity by controlling the motion of the CTD and thus modulating the duplex RNA binding cleft conformation. The interaction of NS3 with NS4B and NS5, being critical for flavivirus replication, can potentially be targeted in therapeutic development against flaviviruses. A complete map of the NS3–NS4B interaction and understanding the mechanism of NS3 helicase activity modulation by the interaction will significantly help the better design of the inhibitory drugs.

(2) The next chapter focused on the NS2B-NS3 two-component protease complex, which will do the majority of the polyprotein processing. A thorough biophysical and biochemical characterization of NS2B3 was done. Invitro-expressed NS2B3 showed a dynamic pH-based oligomerization behavior possibly mediated by the NS2B cofactor region. From all the results from in-vitro experimental and in silico docking and MD simulations studies, the mechanism of how substrate (cleavage

site) recognition happens and how it will reach the catalytic pocket was explained. Like other serine protease surface unstructured loop regions involved in initial substrate recognition, in NS2B3, the part NS2B, which is an conserved electronegative pocket region present on the surface, will be the initial binding pocket for the cleavage site. After this binding, the conserved beta-hairpin motif of NS2B, which fluctuates between the open and closed conformations, will be stabilized in closed conformation while bringing the substrate into the catalytic pocket by handing over the P1 arg/lys of the cleavage site into S1 pocket and second basic residue lys/arg still in contact with the electronegative pocket of NS2B.

(3) In the final chapter, from the initial evidence from NS2B3 biochemical studies, a hypothesis was made for the interaction between NS2B3 and NS4A N-terminal 56 residues region. This chapter did a thorough biophysical characterization of the NS4A protein and established a possible homomeric trimer of NS4A formation mediated NS4A N-terminal 56(NS4AN56) residues region. Then the interaction between NS2B3 and NS4AN56 was established by multiple approaches, and later, it was tested if this NS4AN56 interaction affected the NS2B3 protease activity. From this chapter and taking into consideration of previous studies, it is evident that the last cleavage site of polyprotein protein processing may be the NS3-NS4A cleavage site. After this, NS4A will interact with NS2B3 and negatively regulates its protease activity, which might act as a switch for NS3 protein to switch from polyprotein processing to replication.



NS3 is the central molecule in the overall nonstructural protein interaction network with multiple interactions which are critical in both polyprotein processing and genome replication; in this study, three different nonstructural protein interactions were studied, and their functional significance in the genome replication and polyprotein processing explained in a structural point of few.

Appendices

Primers list

SM5	AGAGCTAGCTCAGGAGTCTTGTGGGAC
SM6	ACTCTCGAGTCTTCTTCCTGCTGCGAA
SM7	AGTGCTAGCTCAGGTGACCTAATATTAG
SM8	TTTCTCGAGTCTGTCTGGCTCTGGAAT
SM10	ACCCTCGAGTCTCCTACCTCCTCCCAA
SM31	CCAGCTAGCGAGATTGAGGACGAGGTG
SM52	CTC GAG TAG TAG CAT CAA TGT TTC TAT
SM59	CAT ATG GGA TCC GCT GAT TTA TCA TTG GAG AGA
SM60	CTC GAG TTT GAG GAG TAT AGT GAG CGT
SM86	ACCACCACCACCAGAACCACCACCACCTTTGAGGAGTA TAGTGAGCGT
SM92	TTC GCA GCA GGA AGA AGA AGT GTC TCA GGT GAC CTA ATA TTA GAA
SM93	TTC TAA TAT TAG GTC ACC TGA GAC ACT TCT TCT TCC TGC TGC GAA
SM96	GGA TCC CAT ATG AAT GAG ATG GGA TTA CTG GAA ACC ACA AAG
SM99	GAT GGA ACT ATG AAA ATA AAA GCT GCA GCG GCA GCT GCC ACG CTC ACT ATA CTC CTC
SM100	GAG GAG TAT AGT GAG CGT GGC AGC TGC CGC TGC AGC TTT TAT TTT CAT AGT TCC ATC
SM144	ACG CCG GGC ACC CGT AAG ACC CCT GAA
SM145	TTC AGG GGT CTT ACG GGT GCC CGG CGT
SM147	GGA TCC CAT ATG GTG AGC AAG GGC GAG GAG
SM148	CGC CGC CAC CAC CAG CGG ATC GCT GCT ATC GTT GCA GTC ACC TGA GAC ACT GCG ACG TCC TGC TGC CTT GTA CAG CTC GTC CAT GCC
SM149	CGC CGC CAC CAC CAG CGG ATC GCT GCT ATC GTT GCA GTC ACC TGA GAC ACT GCG ACG TCC TGC TGC TTT GTA TAG TTC ATC CAT GCC
SM150	TTT CTC GAG CAG ACA ATC CAG AAT CCA CAG AAT CAG ATG CAG AAT GCC GAT GAT GCT CGC CGC CAC CAC CAG CGG ATC
SD2	TTTCATAGTTCCAGCAGCTTGGACCTCCAC
SD1	GTGGAGGTCCAAGCTGCTGGAACTATGAAA

References

1. Oldstone MBA. History of Virology. In: Schmidt TMBT-E of M (Fourth E, ed. Reference Module in Biomedical Sciences. Oxford: Elsevier; 2014:608-612. doi:10.1016/B978-0-12-801238-3.00078-7

2. Sutter PS. “The First Mountain to Be Removed”: Yellow Fever Control and the Construction of the Panama Canal. Environ Hist Durh N C. 2016;21(2):250-259. http://www.ncbi.nlm.nih.gov/pubmed/35309292.

3. Zhang F, Chase-Topping M, Guo C-G, van Bunnik BAD, Brierley L, Woolhouse MEJ. Global discovery of human-infective RNA viruses: A modelling analysis. Morse S, ed. PLoS Pathog. 2020;16(11):e1009079. doi:10.1371/journal.ppat.1009079

4. Carrasco-Hernandez R, Jácome R, López Vidal Y, Ponce de León S. Are RNA Viruses Candidate Agents for the Next Global Pandemic? A Review. ILAR J. 2017;58(3):343-358. doi:10.1093/ilar/ilx026

5. Pierson TC, Diamond MS. The continued threat of emerging flaviviruses. Nat Microbiol. 2020;5(6):796-812. doi:10.1038/s41564-020-0714-0

6. Vasilakis N, Shell EJ, Fokam EB, et al. Potential of ancestral sylvatic dengue-2 viruses to re-emerge. Virology. 2007;358(2):402-412. doi:10.1016/j.virol.2006.08.049

7. Murhekar M V., Kamaraj P, Kumar MS, et al. Burden of dengue infection in India, 2017: a cross-sectional population based serosurvey. Lancet Glob Heal. 2019;7(8):e1065-e1073. doi:10.1016/S2214-109X(19)30250-5

8. Kuhn RJ, Zhang W, Rossmann MG, et al. Structure of Dengue Virus. Cell. 2002;108(5):717-725. doi:10.1016/S0092-8674(02)00660-8

9. World Health Organization & Special Programme for Research and Training in Tropical Diseases. Dengue Guidelines for Diagnosis, Treatment, Prevention and Control. WHO Http://Apps.Who.Int/Iris/ Bitstream/10665/44188/1/9789241547871_ Eng.Pdf (2009).

10. Hadinegoro SR, Arredondo-García JL, Capeding MR, et al. Efficacy and LongTerm Safety of a Dengue Vaccine in Regions of Endemic Disease. N Engl J Med. 2015;373(13):1195-1206. doi:10.1056/nejmoa1506223

11. Kaptein SJF, Goethals O, Kiemel D, et al. A pan-serotype dengue virus inhibitor targeting the NS3–NS4B interaction. Nature. 2021;598(7881):504-509. doi:10.1038/s41586-021-03990-6

12. Noisakran S, Onlamoon N, Songprakhon P, Hsiao H-M, Chokephaibulkit K, Perng GC. Cells in Dengue Virus Infection In Vivo. Adv Virol. 2010;2010:1-15. doi:10.1155/2010/164878

13. Balsitis SJ, Coloma J, Castro G, et al. Tropism of dengue virus in mice and humans defined by viral nonstructural protein 3-specific immunostaining. Am J Trop Med Hyg. 2009;80(3):416-424. http://www.ncbi.nlm.nih.gov/pubmed/19270292.

14. Prestwood TR, May MM, Plummer EM, Morar MM, Yauch LE, Shresta S. Trafficking and Replication Patterns Reveal Splenic Macrophages as Major Targets of Dengue Virus in Mice. J Virol. 2012;86(22):12138-12147. doi:10.1128/JVI.00375-12

15. Surasombatpattana P, Hamel R, Patramool S, et al. Dengue virus replication in infected human keratinocytes leads to activation of antiviral innate immune responses. Infect Genet Evol. 2011;11(7):1664-1673. doi:10.1016/j.meegid.2011.06.009

16. Kongmanas K, Punyadee N, Wasuworawong K, et al. Immortalized stem cellderived hepatocyte-like cells: An alternative model for studying dengue pathogenesis and therapy. Wang W-K, ed. PLoS Negl Trop Dis. 2020;14(11):e0008835. doi:10.1371/journal.pntd.0008835

17. Pham AM, Langlois RA, TenOever BR. Replication in Cells of Hematopoietic Origin Is Necessary for Dengue Virus Dissemination. Kuhn RJ, ed. PLoS Pathog. 2012;8(1):e1002465. doi:10.1371/journal.ppat.1002465

18. Chin JFL, Chu JJH, Ng ML. The envelope glycoprotein domain III of dengue virus serotypes 1 and 2 inhibit virus entry. Microbes Infect. 2007;9(1):1-6. doi:10.1016/j.micinf.2006.09.009

19. Acosta EG, Talarico LB, Damonte EB. Cell entry of dengue virus. Future Virol. 2008;3(5):471-479. doi:10.2217/17460794.3.5.471

20. Oliveira LG, Peron JPS. Viral receptors for flaviviruses: Not only gatekeepers. J Leukoc Biol. 2019;106(3):695-701. doi:10.1002/JLB.MR1118-460R

21. Amara A, Mercer J. Viral apoptotic mimicry. Nat Rev Microbiol. 2015;13(8):461-469. doi:10.1038/nrmicro3469

22. van der Schaar HM, Rust MJ, Chen C, et al. Dissecting the Cell Entry Pathway of Dengue Virus by Single-Particle Tracking in Living Cells. Farzan M, ed. PLoS Pathog. 2008;4(12):e1000244. doi:10.1371/journal.ppat.1000244

23. Acosta EG, Castilla V, Damonte EB. Alternative infectious entry pathways for dengue virus serotypes into mammalian cells. Cell Microbiol. 2009;11(10):1533-1549. doi:10.1111/j.1462-5822.2009.01345.x

24. Westaway EG. FLAVlVlRUS REPLICATION STRATEGY. Adv Virus Res. 1987;33:45-90.

25. Alvarez DE, Lodeiro MF, Luduena SJ, Pietrasanta LI, Gamarnik A V. LongRange RNA-RNA Interactions Circularize the Dengue Virus Genome. J Virol. 2005;79(11):6631-6643. doi:10.1128/jvi.79.11.6631-6643.2005

26. Gebhard LG, Filomatori C V, Gamarnik A V. Functional RNA Elements in the

Dengue Virus Genome. Viruses. 2011;3(9):1739-1756. doi:10.3390/v3091739

27. Villordo SM, Gamarnik A V. Genome cyclization as strategy for flavivirus RNA replication. Virus Res. 2009;139(2):230-239. doi:10.1016/j.virusres.2008.07.016

28. Gritsun TS, Gould EA. Origin and evolution of flavivirus 5′UTRs and panhandles: Trans-terminal duplications? Virology. 2007;366(1):8-15. doi:10.1016/j.virol.2007.04.011

29. Filomatori C V, Lodeiro MF, Alvarez DE, Samsa MM, Pietrasanta L, Gamarnik A V. A 5′ RNA element promotes dengue virus RNA synthesis on a circular genome. Genes Dev. 2006;20(16):2238-2249. doi:10.1101/gad.1444206

30. de Borba L, Villordo SM, Iglesias NG, Filomatori C V, Gebhard LG, Gamarnik A V. Overlapping Local and Long Range RNA-RNA Interactions Modulate Dengue Virus Genome Cyclization and Replication. J Virol. 2015;89(6):JVI.02677--14. doi:10.1128/JVI.02677-14

31. Selisko B, Potisopon S, Agred R, et al. Molecular Basis for Nucleotide Conservation at the Ends of the Dengue Virus Genome. Diamond MS, ed. PLoS Pathog. 2012;8(9):e1002912. doi:10.1371/journal.ppat.1002912

32. Alvarez DE, De Lella Ezcurra AL, Fucito S, Gamarnik A V. Role of RNA structures present at the 3′UTR of dengue virus on translation, RNA synthesis, and viral replication. Virology. 2005;339(2):200-212. doi:10.1016/j.virol.2005.06.009

33. Davis WG, Basu M, Elrod EJ, Germann MW, Brinton MA. Identification of cisActing Nucleotides and a Structural Feature in West Nile Virus 3'-Terminus RNA That Facilitate Viral Minus Strand RNA Synthesis. J Virol. 2013;87(13):7622-7636. doi:10.1128/JVI.00212-13

34. MacKenzie JM, Khromykh AA, Jones MK, Westaway EG. Subcellular localization and some biochemical properties of the flavivirus Kunjin nonstructural proteins NS2A and NS4A. Virology. 1998;245(2):203-215. doi:10.1006/viro.1998.9156

35. Cerikan B, Goellner S, Neufeldt CJ, et al. A Non-Replicative Role of the 3′ Terminal Sequence of the Dengue Virus Genome in Membranous Replication Organelle Formation. Cell Rep. 2020;32(1):107859. doi:10.1016/j.celrep.2020.107859

36. Clyde K, Harris E. RNA Secondary Structure in the Coding Region of Dengue Virus Type 2 Directs Translation Start Codon Selection and Is Required for Viral Replication. J Virol. 2006;80(5):2170-2182. doi:10.1128/JVI.80.5.21702182.2006

37. Dengue Virus Genome Uncoating Requires Ubiquitination. 2016;7(3):1-10. doi:10.1128/mBio.00804-16.Editor

38. Garcia-Blanco MA, Vasudevan SG, Bradrick SS, Nicchitta C. Flavivirus RNA transactions from viral entry to genome replication. Antiviral Res. 2016;134:244249. doi:10.1016/j.antiviral.2016.09.010

39. Holden KL, Harris E. Enhancement of dengue virus translation: role of the 3′ untranslated region and the terminal 3′ stem-loop domain. Virology. 2004;329(1):119-133. doi:10.1016/j.virol.2004.08.004

40. Polacek C, Friebe P, Harris E. Poly(A)-binding protein binds to the nonpolyadenylated 3′ untranslated region of dengue virus and modulates translation efficiency. J Gen Virol. 2009;90(3):687-692. doi:10.1099/vir.0.007021-0

41. Clyde K, Barrera J, Harris E. The capsid-coding region hairpin element (cHP) is a critical determinant of dengue virus and West Nile virus RNA synthesis. Virology. 2008;379(2):314-323. doi:10.1016/j.virol.2008.06.034

42. Edgil D, Polacek C, Harris E. Dengue Virus Utilizes a Novel Strategy for Translation Initiation When Cap-Dependent Translation Is Inhibited. J Virol. 2006;80(6):2976-2986. doi:10.1128/JVI.80.6.2976-2986.2006

43. Tabata K, Arakawa M, Ishida K, et al. Endoplasmic Reticulum-Associated Degradation Controls Virus Protein Homeostasis, Which Is Required for Flavivirus Propagation. James Ou J-H, ed. J Virol. 2021;95(15). doi:10.1128/JVI.02234-20

44. Estoppey D, Lee CM, Janoschke M, et al. The Natural Product Cavinafungin Selectively Interferes with Zika and Dengue Virus Replication by Inhibition of the Host Signal Peptidase. Cell Rep. 2017;19(3):451-460. doi:10.1016/j.celrep.2017.03.071

45. Welsch S, Miller S, Romero-Brey I, et al. Composition and Three-Dimensional Architecture of the Dengue Virus Replication and Assembly Sites. Cell Host Microbe. 2009;5(4):365-375. doi:10.1016/j.chom.2009.03.007

46. Emmott E, de Rougemont A, Hosmillo M, et al. Polyprotein processing and intermolecular interactions within the viral replication complex spatially and temporally control norovirus protease activity. J Biol Chem. 2019;294(11):4259-4271. doi:10.1074/jbc.RA118.006780

47. Shulla A, Randall G. (+) RNA virus replication compartments: A safe home for (most) viral replication. Curr Opin Microbiol. 2016;32:82-88. doi:10.1016/j.mib.2016.05.003

48. den Boon JA, Diaz A, Ahlquist P. Cytoplasmic Viral Replication Complexes. Cell Host Microbe. 2010;8(1):77-85. doi:10.1016/j.chom.2010.06.010

49. Su C-I, Tseng C-H, Yu C-Y, Lai MMC. SUMO Modification Stabilizes Dengue Virus Nonstructural Protein 5 To Support Virus Replication. Jung JU, ed. J Virol. 2016;90(9):4308-4319. doi:10.1128/JVI.00223-16

50. Junjhon J, Pennington JG, Edwards TJ, Perera R, Lanman J, Kuhn RJ. Ultrastructural Characterization and Three-Dimensional Architecture of Replication Sites in Dengue Virus-Infected Mosquito Cells. J Virol. 2014;88(9):4687-4697. doi:10.1128/JVI.00118-14

51. Chatel-Chaix L, Cortese M, Romero-Brey I, et al. Dengue Virus Perturbs Mitochondrial Morphodynamics to Dampen Innate Immune Responses. Cell Host Microbe. 2016;20(3):342-356. doi:10.1016/j.chom.2016.07.008
52. Uchil PD, Satchidanandam V. Architecture of the Flaviviral Replication Complex. J Biol Chem. 2003;278(27):24388-24398. doi:10.1074/jbc.M301717200
53. Cortese M, Goellner S, Acosta EG, et al. Ultrastructural Characterization of Zika Virus Replication Factories. Cell Rep. 2017;18(9):2113-2123. doi:10.1016/j.celrep.2017.02.014
54. Offerdahl DK, Dorward DW, Hansen BT, Bloom ME. A Three-Dimensional Comparison of Tick-Borne Flavivirus Infection in Mammalian and Tick Cell Lines. Ganta R, ed. PLoS One. 2012;7(10):e47912. doi:10.1371/journal.pone.0047912
55. Gillespie LK, Hoenen A, Morgan G, Mackenzie JM. The Endoplasmic Reticulum Provides the Membrane Platform for Biogenesis of the Flavivirus Replication Complex. J Virol. 2010;84(20):10438-10447. doi:10.1128/JVI.00986-10
56. McLean JE, Wudzinska A, Datan E, Quaglino D, Zakeri Z. Flavivirus NS4Ainduced Autophagy Protects Cells against Death and Enhances Virus Replication. J Biol Chem. 2011;286(25):22147-22159. doi:10.1074/jbc.M110.192500
57. Suhy DA, Giddings TH, Kirkegaard K. Remodeling the Endoplasmic Reticulum by Poliovirus Infection and by Individual Viral Proteins: an Autophagy-Like Origin for Virus-Induced Vesicles. J Virol. 2000;74(19):8953-8965. doi:10.1128/JVI.74.19.8953-8965.2000
58. Heaton NS, Perera R, Berger KL, et al. Dengue virus nonstructural protein 3 redistributes fatty acid synthase to sites of viral replication and increases cellular fatty acid synthesis. Proc Natl Acad Sci. 2010;107(40):17345-17350. doi:10.1073/pnas.1010811107
59. Welsch S, Miller S, Romero-Brey I, et al. Composition and Three-Dimensional Architecture of the Dengue Virus Replication and Assembly Sites. Cell Host Microbe. 2009;5(4):365-375. doi:10.1016/j.chom.2009.03.007
60. Raviprakash K, Porter KR, Hayes CG, Sinha M. Conversion of dengue virus replicative form RNA (RF) to replicative intermediate (RI) by nonstructural proteins NS-5 and NS-3. Am J Trop Med Hyg. 1998;58(1):90-95. doi:10.4269/ajtmh.1998.58.90
61. Klema V, Padmanabhan R, Choi K. Flaviviral Replication Complex: Coordination between RNA Synthesis and 5'-RNA Capping. Viruses. 2015;7(8):4640-4656. doi:10.3390/v7082837
62. Cleaves GR, Ryan TE, Walter Schlesinger R. Identification and characterization of type 2 dengue virus replicative intermediate and replicative form RNAs. Virology. 1981;111(1):73-83. doi:10.1016/0042-6822(81)90654-1

63. Lindenbach BD, Rice CM. trans-Complementation of yellow fever virus NS1 reveals a role in early RNA replication. J Virol. 1997;71(12):9608-9617. doi:10.1128/JVI.71.12.9608-9617.1997

64. Westaway EG, Mackenzie JM, Kenney MT, Jones MK, Khromykh AA. Ultrastructure of Kunjin virus-infected cells: colocalization of NS1 and NS3 with double-stranded RNA, and of NS2B with NS3, in virus-induced membrane structures. J Virol. 1997;71(9):6650-6661. doi:10.1128/JVI.71.9.6650-6661.1997

65. MACKENZIE JM, JONES MK, YOUNG PR. Immunolocalization of the Dengue Virus Nonstructural Glycoprotein NS1 Suggests a Role in Viral RNA Replication. Virology. 1996;220(1):232-240. doi:10.1006/viro.1996.0307

66. Cervantes-Salazar M, Angel-Ambrocio AH, Soto-Acosta R, et al. Dengue virus NS1 protein interacts with the ribosomal protein RPL18: This interaction is required for viral translation and replication in Huh-7 cells. Virology. 2015;484:113-126. doi:10.1016/j.virol.2015.05.017

67. Westaway EG, Goodman MR. Variation in distribution of the three flavivirusspecified glycoproteins detected by immunofluorescence in infected Vero cells. Arch Virol. 1987;94(3-4):215-228. doi:10.1007/BF01310715

68. Akey DL, Clay Brown W, Dutta S, et al. Flavivirus NS1 crystal structures reveal a surface for membrane association and regions of interaction with the immune system. Science (80-). 2014;343(6173):881-885. doi:10.1126/science.1247749.Flavivirus

69. Winkler G, Maxwell SE, Ruemmler C, Stollar V. Newly synthesized dengue-2 virus nonstructural protein NS1 is a soluble protein but becomes partially hydrophobic and membrane-associated after dimerization. Virology. 1989;171(1):302-305. doi:10.1016/0042-6822(89)90544-8

70. Falgout B, Chanock R, Lai CJ. Proper processing of dengue virus nonstructural glycoprotein NS1 requires the N-terminal hydrophobic signal sequence and the downstream nonstructural protein NS2a. J Virol. 1989;63(5):1852-1860. doi:10.1128/jvi.63.5.1852-1860.1989

71. Xu X, Song H, Qi J, et al. Contribution of intertwined loop to membrane association revealed by Zika virus full-length <scp>NS</scp> 1 structure. EMBO J. 2016;35(20):2170-2178. doi:10.15252/embj.201695290

72. Ci Y, Liu Z-Y, Zhang N-N, et al. Zika NS1–induced ER remodeling is essential for viral replication. J Cell Biol. 2020;219(2). doi:10.1083/jcb.201903062

73. Preugschat F, Yao CW, Strauss JH. In vitro processing of dengue virus type 2 nonstructural proteins NS2A, NS2B, and NS3. J Virol. 1990;64(9):4364-4374. doi:10.1128/JVI.64.9.4364-4374.1990

74. Falgout B, Markoff L. Evidence that flavivirus NS1-NS2A cleavage is mediated by a membrane-bound host protease in the endoplasmic reticulum. J Virol. 1995;69(11):7232-7243. doi:10.1128/JVI.69.11.7232-7243.1995

75. Xie X, Gayen S, Kang C, Yuan Z, Shi P-Y. Membrane Topology and Function of Dengue Virus NS2A Protein. J Virol. 2013;87(8):4609-4622. doi:10.1128/jvi.02424-12

76. Leung JY, Pijlman GP, Kondratieva N, Hyde J, Mackenzie JM, Khromykh AA. Role of Nonstructural Protein NS2A in Flavivirus Assembly. J Virol. 2008;82(10):4731-4741. doi:10.1128/jvi.00002-08

77. Xie X, Zou J, Puttikhunt C, Yuan Z, Shi P-Y. Two Distinct Sets of NS2A Molecules Are Responsible for Dengue Virus RNA Synthesis and Virion Assembly. J Virol. 2015;89(2):1298-1313. doi:10.1128/JVI.02882-14

78. Melian EB, Edmonds JH, Nagasaki TK, Hinzman E, Floden N, Khromykh AA. West Nile virus NS2A protein facilitates virus-induced apoptosis independently of interferon response. J Gen Virol. 2013;94(PART2):308-313. doi:10.1099/vir.0.047076-0

79. Ye Q, Li XF, Zhao H, et al. A single nucleotide mutation in NS2A of Japanese encephalitis-live vaccine virus (SA14-14-2) ablates NS1' formation and contributes to attenuation. J Gen Virol. 2012;93(PART 9):1959-1964. doi:10.1099/vir.0.043844-0

80. Liu WJ, Chen HB, Wang XJ, Huang H, Khromykh AA. Analysis of Adaptive Mutations in Kunjin Virus Replicon RNA Reveals a Novel Role for the Flavivirus Nonstructural Protein NS2A in Inhibition of Beta Interferon PromoterDriven Transcription. J Virol. 2004;78(22):12225-12235. doi:10.1128/jvi.78.22.12225-12235.2004

81. Speight G, Coia G, Parker MD, Westaway EG. Gene mapping and positive identification of the non-structural proteins NS2A, NS2B, NS3, NS4B and NS5 of the flavivirus Kunjin and their cleavage sites. J Gen Virol. 1988;69(1):23-34. doi:10.1099/0022-1317-69-1-23

82. Clum S, Ebner KE, Padmanabhan R. Cotranslational membrane insertion of the serine proteinase precursor NS2B-NS3(Pro) of dengue virus type 2 is required for efficient in vitro processing and is mediated through the hydrophobic regions of NS2B. J Biol Chem. 1997;272(49):30715-30723. doi:10.1074/jbc.272.49.30715

83. Arias CF, Preugschat F, Strass JH. Dengue 2 virus ns2b and ns3 form a stable complex that can cleave ns3 within the helicase domain. Virology. 1993;193(2):888-899. doi:10.1006/viro.1993.1198

84. Falgout B, Miller RH, Lai CJ. Deletion analysis of dengue virus type 4 nonstructural protein NS2B: identification of a domain required for NS2B-NS3 protease activity. J Virol. 1993;67(4):2034-2042. doi:10.1128/jvi.67.4.20342042.1993

85. Yusof R, Clum S, Wetzel M, Murthy HMK, Padmanabhan R. Purified NS2B/NS3 serine protease of dengue virus type 2 exhibits cofactor NS2B dependence for cleavage of substrates with dibasic amino acids in vitro. J Biol Chem. 2000;275(14):9963-9969. doi:10.1074/JBC.275.14.9963

86. Niyomrattanakit P, Winoyanuwattikun P, Chanprapaph S, Angsuthanasombat C, Panyim S, Katzenmeier G. Identification of Residues in the Dengue Virus Type 2 NS2B Cofactor That Are Critical for NS3 Protease Activation. J Virol. 2004;78(24):13708-13716. doi:10.1128/JVI.78.24.13708-13716.2004

87. Gupta G, Lim L, Song J. NMR and MD studies reveal that the isolated dengue NS3 protease is an intrinsically disordered chymotrypsin fold which absolutely requests NS2B for correct folding and functional dynamics. PLoS One. 2015;10(8):1-24. doi:10.1371/journal.pone.0134823

88. Xing H, Xu S, Jia F, et al. Zika NS2B is a crucial factor recruiting NS3 to the ER and activating its protease activity. Virus Res. 2020;275:197793. doi:10.1016/j.virusres.2019.197793

89. León-Juárez M, Martínez-Castillo M, Shrivastava G, et al. Recombinant Dengue virus protein NS2B alters membrane permeability in different membrane models. Virol J. 2016;13(1):1. doi:10.1186/s12985-015-0456-4

90. Li X, Deng C, Ye H, et al. Transmembrane Domains of NS2B Contribute to both Viral RNA. 2016;90(12):5735-5749. doi:10.1128/JVI.00340-16.Editor

91. Yaegashi T, Vakharia VN, Page K, Sasaguri Y, Feighny R, Padmanabhan R. Partial sequence analysis of cloned dengue virus type 2 genome. Gene. 1986;46(2-3):257-267. doi:10.1016/0378-1119(86)90410-5

92. Mackow E, Makino Y, Zhao B, et al. The nucleotide sequence of dengue type 4 virus: Analysis of genes coding for nonstructural proteins. Virology. 1987;159(2):217-228. doi:10.1016/0042-6822(87)90458-2

93. Zhang L, Mohan PM, Padmanabhan R. Processing and localization of Dengue virus type 2 polyprotein precursor NS3-NS4A-NS4B-NS5. J Virol. 1992;66(12):7549-7554. doi:10.1128/jvi.66.12.7549-7554.1992

94. Lai C-J, Pethel M, Jan LR, Kawano H, Cahour A, Falgout B. Processing of dengue type 4 and other flavivirus nonstructural proteins. In: Positive-Strand RNA Viruses. Vienna: Springer Vienna; 1994:359-368. doi:10.1007/978-3-70919326-6_36

95. Xu T, Sampath A, Chao A, et al. Structure of the Dengue Virus Helicase/Nucleoside Triphosphatase Catalytic Domain at a Resolution of 2.4 Å. J Virol. 2005;79(16):10278-10288. doi:10.1128/JVI.79.16.10278-10288.2005

96. Chambers TJ, Weir RC, Grakoui A, et al. Evidence that the N-terminal domain of nonstructural protein NS3 from yellow fever virus is a serine protease responsible for site-specific cleavages in the viral polyprotein. Proc Natl Acad Sci U S A. 1990;87(22):8898-8902. doi:10.1073/pnas.87.22.8898

97. Chernov A V, Shiryaev SA, Aleshin AE, et al. The Two-component NS2B-NS3 Proteinase Represses DNA Unwinding Activity of the West Nile Virus NS3 Helicase. J Biol Chem. 2008;283(25):17270-17278. doi:10.1074/jbc.M801719200

98. Lee E, Fernon C, Simpson R, Weir R, Rice C, Dalgarno L. Sequence of the 3?

half of the Murray Valley encephalitis virus genome and mapping of the nonstructural proteins NS1, NS3, and NS5. Virus Genes. 1990;4(3). doi:10.1007/BF00265630

99. Gorbalenya AE, Koonin E V, Donchenko AP, Blinov VM. A novel superfamily of nucleoside triphosphate-binding motif containing proteins which are probably involved in duplex unwinding in DNA and RNA replication and recombination. FEBS Lett. 1988;235(1-2):16-24. doi:10.1016/0014-5793(88)81226-2

100. Byrd AK, Raney KD. Superfamily 2 helicases. Front Biosci. 2012;17(6):20702088. doi:10.2741/4038

101. Li H, Clum S, You S, Ebner KE, Padmanabhan R. 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. 1999;73(4):3108-3116. doi:10.1128/JVI.73.4.3108-3116.1999

102. Pyle AM. Translocation and Unwinding Mechanisms of RNA and DNA Helicases. Annu Rev Biophys. 2008;37(1):317-336. doi:10.1146/annurev.biophys.37.032807.125908

103. Benarroch D, Selisko B, Locatelli GA, Maga G, Romette JL, Canard B. The RNA helicase, nucleotide 5′-triphosphatase, and RNA 5′-triphosphatase activities of Dengue virus protein NS3 are Mg 2+-dependent and require a functional Walker B motif in the helicase catalytic core. Virology. 2004;328(2):208-218. doi:10.1016/J.VIROL.2004.07.004

104. Du Pont KE, Davidson RB, McCullagh M, Geiss BJ. Motif V regulates energy transduction between the flavivirus NS3 ATPase and RNA-binding cleft. J Biol Chem. 2020;295(6):1551-1564. doi:10.1074/jbc.RA119.011922

105. Davidson RB, Hendrix J, Geiss BJ, McCullagh M. Allostery in the dengue virus NS3 helicase: Insights into the NTPase cycle from molecular simulations. PLoS Comput Biol. 2018;14(4):1-28. doi:10.1371/journal.pcbi.1006103

106. Swarbrick CMD, Basavannacharya C, Chan KWK, et al. NS3 helicase from dengue virus specifically recognizes viral RNA sequence to ensure optimal replication. Nucleic Acids Res. 2017;45(22):12904-12920. doi:10.1093/nar/gkx1127

107. Luo D, Xu T, Watson RP, et al. Insights into RNA unwinding and ATP hydrolysis by the flavivirus NS3 protein. EMBO J. 2008;27(23):3209-3219. doi:10.1038/emboj.2008.232

108. Gebhard LG, Kaufman SB, Gamarnik A V. Novel ATP-Independent RNA Annealing Activity of the Dengue Virus NS3 Helicase. Marcello A, ed. PLoS One. 2012;7(4):e36244. doi:10.1371/journal.pone.0036244

109. Gebhard LG, Iglesias NG, Byk LA, Filomatori C V, De Maio FA, Gamarnik A V. A Proline-Rich N-Terminal Region of the Dengue Virus NS3 Is Crucial for Infectious Particle Production. Diamond MS, ed. J Virol. 2016;90(11):54515461. doi:10.1128/JVI.00206-16

110. Speight G, Westaway EG. Positive identification of NS4A, the last of the hypothetical nonstructural proteins of flaviviruses. Virology. 1989;170(1):299-301. doi:10.1016/0042-6822(89)90383-8

111. Roosendaal J, Westaway EG, Khromykh A, Mackenzie JM. Regulated Cleavages at the West Nile Virus NS4A-2K-NS4B Junctions Play a Major Role in Rearranging Cytoplasmic Membranes and Golgi Trafficking of the NS4A Protein. J Virol. 2006;80(9):4623-4632. doi:10.1128/JVI.80.9.4623-4632.2006

112. Li Y, Lee MY, Loh YR, Kang CB. Secondary structure and membrane topology of dengue virus NS4A protein in micelles. Biochim Biophys Acta - Biomembr. 2018;1860(2):442-450. doi:10.1016/j.bbamem.2017.10.016

113. Miller S, Kastner S, Krijnse-Locker J, Bühler S, Bartenschlager R. The nonstructural protein 4A of dengue virus is an integral membrane protein inducing membrane alterations in a 2K-regulated manner. J Biol Chem. 2007;282(12):8873-8882. doi:10.1074/jbc.M609919200

114. Hung Y-F, Schwarten M, Schünke S, et al. Dengue virus NS4A cytoplasmic domain binding to liposomes is sensitive to membrane curvature. Biochim Biophys Acta - Biomembr. 2015;1848(5):1119-1126. doi:10.1016/j.bbamem.2015.01.015

115. Lee CM, Xie X, Zou J, et al. Determinants of Dengue Virus NS4A Protein Oligomerization. J Virol. 2015;89(12):6171-6183. doi:10.1128/jvi.00546-15

116. Ambrose RL, Mackenzie JM. Conserved amino acids within the N-terminus of the West Nile virus NS4A protein contribute to virus replication, protein stability and membrane proliferation. Virology. 2015;481:95-106. doi:10.1016/j.virol.2015.02.045

117. Stern O, Hung Y-F, Valdau O, et al. An N-Terminal Amphipathic Helix in Dengue Virus Nonstructural Protein 4A Mediates Oligomerization and Is Essential for Replication. J Virol. 2013;87(7):4080-4085. doi:10.1128/jvi.0190012

118. Lin C-W, Cheng C-W, Yang T-C, et al. Interferon antagonist function of Japanese encephalitis virus NS4A and its interaction with DEAD-box RNA helicase DDX42. Virus Res. 2008;137(1):49-55. doi:10.1016/j.virusres.2008.05.015

119. Dalrymple NA, Cimica V, Mackow ER. Dengue Virus NS Proteins Inhibit RIGI/MAVS Signaling by Blocking TBK1/IRF3 Phosphorylation: Dengue Virus Serotype 1 NS4A Is a Unique Interferon-Regulating Virulence Determinant. Buchmeier MJ, ed. MBio. 2015;6(3). doi:10.1128/mBio.00553-15

120. Cahour A, Falgout B, Lai CJ. Cleavage of the dengue virus polyprotein at the NS3/NS4A and NS4B/NS5 junctions is mediated by viral protease NS2B-NS3, whereas NS4A/NS4B may be processed by a cellular protease. J Virol. 1992;66(3):1535-1542. doi:10.1128/jvi.66.3.1535-1542.1992

121. Płaszczyca A, Scaturro P, Neufeldt CJ, et al. A novel interaction between dengue virus nonstructural protein 1 and the NS4A-2K-4B precursor is required for viral RNA replication but not for formation of the membranous replication organelle.

Randall G, ed. PLOS Pathog. 2019;15(5):e1007736. doi:10.1371/journal.ppat.1007736

122. Zmurko J, Neyts J, Dallmeier K. Flaviviral NS4b, chameleon and jack-in-the-box roles in viral replication and pathogenesis, and a molecular target for antiviral intervention. Rev Med Virol. 2015;25(4):205-223. doi:10.1002/rmv.1835

123. Tian J-N, Yang C-C, Chuang C-K, et al. A Dengue Virus Type 2 (DENV-2) NS4B-Interacting Host Factor, SERP1, Reduces DENV-2 Production by Suppressing Viral RNA Replication. Viruses. 2019;11(9):787. doi:10.3390/v11090787

124. Xie X, Zou J, Wang Q-Y, Shi P-Y. Targeting dengue virus NS4B protein for drug discovery. Antiviral Res. 2015;118:39-45. doi:10.1016/j.antiviral.2015.03.007

125. Wu Y-W, Mettling C, Wu S-R, et al. Autophagy-associated dengue vesicles promote viral transmission avoiding antibody neutralization. Sci Rep. 2016;6(1):32243. doi:10.1038/srep32243

126. Naik NG, Wu H-N. Mutation of Putative N-Glycosylation Sites on Dengue Virus NS4B Decreases RNA Replication. J Virol. 2015;89(13):6746-6760. doi:10.1128/jvi.00423-15

127. Zou J, Xie X, Lee LT, et al. Dimerization of Flavivirus NS4B Protein. J Virol. 2014;88(6):3379-3391. doi:10.1128/jvi.02782-13

128. Zhou Y, Ray D, Zhao Y, et al. Structure and Function of Flavivirus NS5 Methyltransferase. J Virol. 2007;81(8):3891-3903. doi:10.1128/JVI.02704-06

129. Issur M, Geiss BJ, Bougie I, et al. The flavivirus NS5 protein is a true RNA guanylyltransferase that catalyzes a two-step reaction to form the RNA cap structure. RNA. 2009;15(12):2340-2350. doi:10.1261/rna.1609709

130. Potisopon S, Priet S, Collet A, Decroly E, Canard B, Selisko B. The methyltransferase domain of dengue virus protein NS5 ensures efficient RNA synthesis initiation and elongation by the polymerase domain. Nucleic Acids Res. 2014;42(18):11642-11656. doi:10.1093/nar/gku666

131. Li A, Wang W, Wang Y, et al. NS5 Conservative Site Is Required for Zika Virus to Restrict the RIG-I Signaling. Front Immunol. 2020;11. doi:10.3389/fimmu.2020.00051

132. Klema VJ, Ye M, Hindupur A, et al. Dengue Virus Nonstructural Protein 5 (NS5) Assembles into a Dimer with a Unique Methyltransferase and Polymerase Interface. Rey FA, ed. PLOS Pathog. 2016;12(2):e1005451. doi:10.1371/journal.ppat.1005451

133. Bartholomeusz AI, Wright PJ. Synthesis of dengue virus RNA in vitro: initiation and the involvement of proteins NS3 and NS5. Arch Virol. 1993;128(1-2):111121. doi:10.1007/BF01309792

134. Selisko B, Dutartre H, Guillemot J-C, et al. Comparative mechanistic studies of de novo RNA synthesis by flavivirus RNA-dependent RNA polymerases. Virology. 2006;351(1):145-158. doi:10.1016/j.virol.2006.03.026

135. Ji W, Luo G. Zika virus NS5 nuclear accumulation is protective of protein degradation and is required for viral RNA replication. Virology. 2020;541:124-135. doi:10.1016/j.virol.2019.10.010
136. Pryor MJ, Rawlinson SM, Butcher RE, et al. Nuclear Localization of Dengue Virus Nonstructural Protein 5 Through Its Importin $α$/$β$-Recognized Nuclear Localization Sequences is Integral to Viral Infection. Traffic. 2007;8(7):795-807. doi:10.1111/j.1600-0854.2007.00579.x
137. Dengue Virus Non-Structural Protein 5. Viruses. 2017;9(4):91. doi:10.3390/v9040091
138. Lopez-Denman A, Mackenzie J. The IMPORTance of the Nucleus during Flavivirus Replication. Viruses. 2017;9(1):14. doi:10.3390/v9010014
139. Xie X, Zou J, Zhang X, et al. Dengue NS2A Protein Orchestrates Virus Assembly. Cell Host Microbe. 2019;26(5):606--622.e8. doi:10.1016/j.chom.2019.09.015
140. Shiryaev SA, Chernov A V, Aleshin AE, Shiryaeva TN, Strongin AY. NS4A regulates the ATPase activity of the NS3 helicase: a novel cofactor role of the non-structural protein NS4A from West Nile virus. J Gen Virol. 2009;90(9):2081-2085. doi:10.1099/vir.0.012864-0
141. Youn S, Li T, McCune BT, et al. Evidence for a Genetic and Physical Interaction between Nonstructural Proteins NS1 and NS4B That Modulates Replication of West Nile Virus. J Virol. 2012;86(13):7360-7371. doi:10.1128/JVI.00157-12
142. Hafirassou ML, Meertens L, Umaña-Diaz C, et al. A Global Interactome Map of the Dengue Virus NS1 Identifies Virus Restriction and Dependency Host Factors. Cell Rep. 2017;21(13):3900-3913. doi:10.1016/j.celrep.2017.11.094
143. Zou J, Xie X, Wang Q-Y, et al. Characterization of Dengue Virus NS4A and NS4B Protein Interaction. Dermody TS, ed. J Virol. 2015;89(7):3455-3470. doi:10.1128/JVI.03453-14
144. Voßmann S, Wieseler J, Kerber R, Kümmerer BM. A Basic Cluster in the N Terminus of Yellow Fever Virus NS2A Contributes to Infectious Particle Production. J Virol. 2015;89(9):4951-4965. doi:10.1128/jvi.03351-14
145. Xu S, Ci Y, Wang L, et al. Zika virus NS3 is a canonical RNA helicase stimulated by NS5 RNA polymerase. Nucleic Acids Res. 2019;47(16):8693-8707. doi:10.1093/nar/gkz650
146. Tay MYF, Vasudevan SG. The Transactions of NS3 and NS5 in Flaviviral RNA Replication. In: Advances in Experimental Medicine and Biology. Vol 1062. United States; 2018:147-163. doi:10.1007/978-981-10-8727-1_11
147. Tay MYF, Saw WG, Zhao Y, et al. The C-terminal 50 Amino Acid Residues of Dengue NS3 Protein Are Important for NS3-NS5 Interaction and Viral Replication. J Biol Chem. 2015;290(4):2379-2394. doi:10.1074/jbc.M114.607341

148. Zou J, Lee LT, Wang QY, et al. Mapping the Interactions between the NS4B and NS3 Proteins of Dengue Virus. Dermody TS, ed. J Virol. 2015;89(7):3471-3483. doi:10.1128/JVI.03454-14

149. Umareddy I, Chao A, Sampath A, Gu F, Vasudevan SG. Dengue virus NS4B interacts with NS3 and dissociates it from single-stranded RNA. J Gen Virol. 2006;87(9):2605-2614. doi:10.1099/vir.0.81844-0

150. Teramoto T, Balasubramanian A, Choi KH, Padmanabhan R. Serotype-specific interactions among functional domains of dengue virus 2 nonstructural proteins (NS) 5 and NS3 are crucial for viral RNA replication. J Biol Chem. 2017;292(23):9465-9479. doi:10.1074/jbc.M117.775643

151. Chandramouli S, Joseph JS, Daudenarde S, Gatchalian J, Cornillez-Ty C, Kuhn P. Serotype-Specific Structural Differences in the Protease-Cofactor Complexes of the Dengue Virus Family. J Virol. 2010;84(6):3059-3067. doi:10.1128/JVI.02044-09

152. Champreda V, Khumthong R, Subsin B, Angsuthanasombat C, Panyim S, Katzenmeier G. The Two-Component Protease NS2B-NS3 of Dengue Virus Type 2: Cloning, Expression in Escherichia coli and Purification of the NS2B, NS3(pro) and NS2B-NS3 Proteins. J Biochem Mol Biol. 2000;33(4):294-299.

153. Chappell KJ, Stoermer MJ, Fairlie DP, Young PR. Insights to Substrate Binding and Processing by West Nile Virus NS3 Protease through Combined Modeling, Protease Mutagenesis, and Kinetic Studies. J Biol Chem. 2006;281(50):38448-38458. doi:10.1074/jbc.M607641200

154. Urbani A, Bianchi E, Narjes F, et al. Substrate specificity of the hepatitis C virus serine protease NS3. J Biol Chem. 1997;272(14):9204-9209. doi:10.1074/jbc.272.14.9204

155. Valle RP, Falgout B. Mutagenesis of the NS3 protease of dengue virus type 2. J Virol. 1998;72(1):624-632. doi:10.1128/JVI.72.1.624-632.1998

156. Czapinska H, Otlewski J. Structural and energetic determinants of the S1-site specificity in serine proteases. Eur J Biochem. 1999;260(3):571-595. doi:10.1046/j.1432-1327.1999.00160.x

157. Cera E Di. Serine Proteases Enrico. Int Union Biochem Mol Biol Life. 2009;61(5):510-515. doi:10.1002/iub.186.Serine

158. Kumar D, Kumar A, Bhardwaj T, Giri R. Zika virus NS4A N-Terminal region (1-48) acts as a cofactor for inducing NTPase activity of NS3 helicase but not NS3 protease. Arch Biochem Biophys. 2020;695:108631. doi:10.1016/j.abb.2020.108631

159. Neufeldt CJ, Cortese M, Acosta EG, Bartenschlager R. Rewiring cellular networks by members of the Flaviviridae family. Nat Rev Microbiol. 2018;16(3):125-142. doi:10.1038/nrmicro.2017.170

160. van den Elsen K, Quek JP, Luo D. Molecular Insights into the Flavivirus Replication Complex. Viruses. 2021;13(6):956. doi:10.3390/v13060956

161. Chatel-Chaix L, Fischl W, Scaturro P, et al. A Combined Genetic-Proteomic Approach Identifies Residues within Dengue Virus NS4B Critical for Interaction with NS3 and Viral Replication. J Virol. 2015;89(14):7170-7186. doi:10.1128/JVI.00867-15

162. Xie X, Wang Q-Y, Xu HY, et al. Inhibition of Dengue Virus by Targeting Viral NS4B Protein. J Virol. 2011;85(21):11183-11195. doi:10.1128/JVI.05468-11

163. Luo D, Wei N, Doan DN, et al. Flexibility between the protease and helicase domains of the dengue virus NS3 protein conferred by the linker region and its functional implications. J Biol Chem. 2010;285(24):18817-18827. doi:10.1074/jbc.M109.090936

164. Luo D, Xu T, Hunke C, Grüber G, Vasudevan SG, Lescar J. Crystal Structure of the NS3 Protease-Helicase from Dengue Virus. J Virol. 2008;82(1):173-183. doi:10.1128/jvi.01788-07

165. Donsbach P, Klostermeier D. Regulation of RNA helicase activity: principles and examples. Biol Chem. 2021;402(5):529-559. doi:10.1515/hsz-2020-0362

166. Yokota H. Roles of the C-Terminal Amino Acids of Non-Hexameric Helicases: Insights from Escherichia coli UvrD. Int J Mol Sci. 2021;22(3):1018. doi:10.3390/ijms22031018

167. Miller S, Sparacio S, Bartenschlager R. Subcellular Localization and Membrane Topology of the Dengue Virus Type 2 Non-structural Protein 4B * □ S. 2006. doi:10.1074/jbc.M512697200

168. Lu H, Zhan Y, Li X, et al. Novel insights into the function of an N-terminal region of DENV2 NS4B for the optimal helicase activity of NS3. Virus Res. 2021;295:198318. doi:https://doi.org/10.1016/j.virusres.2021.198318

169. Muñoz-Jordán JL, Laurent-Rolle M, Ashour J, et al. Inhibition of Alpha/Beta Interferon Signaling by the NS4B Protein of Flaviviruses. J Virol. 2005;79(13):8004-8013. doi:10.1128/JVI.79.13.8004-8013.2005

170. Jain R, Coloma J, García-Sastre A, Aggarwal AK. Structure of the NS3 helicase from Zika virus. Nat Struct Mol Biol. 2016;23(8):752-754. doi:10.1038/nsmb.3258

171. van Cleef KWR, Overheul GJ, Thomassen MC, et al. Identification of a new dengue virus inhibitor that targets the viral NS4B protein and restricts genomic RNA replication. Antiviral Res. 2013;99(2):165-171. doi:10.1016/j.antiviral.2013.05.011

172. Karimova G, Pidoux J, Ullmann A, Ladant D. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. Proc Natl Acad Sci. 1998;95(10):5752-5756. doi:10.1073/pnas.95.10.5752

173. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-589. doi:10.1038/s41586-02103819-2

174. Baek M, DiMaio F, Anishchenko I, et al. Accurate prediction of protein structures and interactions using a three-track neural network. Science (80-). 2021;373(6557):871-876. doi:10.1126/science.abj8754

175. Dominguez C, Boelens R, Bonvin AMJJ. HADDOCK: A Protein−Protein Docking Approach Based on Biochemical or Biophysical Information. J Am Chem Soc. 2003;125(7):1731-1737. doi:10.1021/ja026939x

176. Kozakov D, Hall DR, Xia B, et al. The ClusPro web server for protein–protein docking. Nat Protoc. 2017;12(2):255-278. doi:10.1038/nprot.2016.169

177. Greenfield NJ. Circular Dichroism (CD) Analyses of Protein-Protein Interactions. In: Meyerkord CL, Fu H, eds. Protein-Protein Interactions. Methods in Molecular Biology. New York, NY: Springer New York; 2015:239-265. doi:10.1007/978-1-4939-2425-7_15

178. Phizicky EM, Fields S. Protein-protein interactions: methods for detection and analysis. Microbiol Rev. 1995;59(1):94-123. doi:10.1128/mr.59.1.94-123.1995

179. Hoffmann H-H, Schneider WM, Rozen-Gagnon K, et al. TMEM41B Is a Panflavivirus Host Factor. Cell. 2021;184(1):133-148.e20. doi:10.1016/j.cell.2020.12.005

180. Tian H, Ji X, Yang X, et al. Structural basis of Zika virus helicase in recognizing its substrates. Protein Cell. 2016;7(8):562-570. doi:10.1007/s13238-016-0293-2

181. Lescar J, Soh S, Lee LT, Vasudevan SG, Kang C, Lim SP. The Dengue Virus Replication Complex: From RNA Replication to Protein-Protein Interactions to Evasion of Innate Immunity. In: ; 2018:115-129. doi:10.1007/978-981-10-87271_9

182. Johansson M, Brooks AJ, Jans DA, Vasudevan SG. A small region of the dengue virus-encoded RNA-dependent RNA polymerase, NS5, confers interaction with both the nuclear transport receptor importin-β and the viral helicase, NS3. J Gen Virol. 2001;82(4):735-745. doi:10.1099/0022-1317-82-4-735

183. Miller S, Sparacio S, Bartenschlager R. Subcellular Localization and Membrane Topology of the Dengue Virus Type 2 Non-structural Protein 4B. J Biol Chem. 2006;281(13):8854-8863. doi:10.1074/jbc.M512697200

184. Zheng W, Liao J-C, Brooks BR, Doniach S. Toward the mechanism of dynamical couplings and translocation in hepatitis C virus NS3 helicase using elastic network model. Proteins Struct Funct Bioinforma. 2007;67(4):886-896. doi:10.1002/prot.21326

185. Davidson RB, Hendrix J, Geiss BJ, Mccullagh M. Allostery in the dengue virus NS3 helicase: Insights into the NTPase cycle from molecular simulations. PLoS Comput Biol. 2018;14(4):e1006103. doi:10.1371/journal.pcbi.1006103

186. Yamshchikov VF, Compans RW. Formation of the flavivirus envelope: role of the viral NS2B-NS3 protease. J Virol. 1995;69(4):1995-2003. doi:10.1128/jvi.69.4.1995-2003.1995

187. Lin C, Chambers TJ, Rice CM. Mutagenesis of Conserved Residues at the Yellow Fever Virus 3/4A and 4B/5 Dibasic Cleavage Sites: Effects on Cleavage Efficiency and Polyprotein Processing. Virology. 1993;192(2):596-604. doi:10.1006/viro.1993.1076

188. Amberg SM, Nestorowicz A, McCourt DW, Rice CM. NS2B-3 proteinasemediated processing in the yellow fever virus structural region: in vitro and in
vivo studies. J Virol. 1994;68(6):3794-3802. doi:10.1128/jvi.68.6.37943802.1994

189. Lin C, Amberg SM, Chambers TJ, Rice CM. Cleavage at a novel site in the NS4A region by the yellow fever virus NS2B-3 proteinase is a prerequisite for processing at the downstream 4A/4B signalase site. J Virol. 1993;67(4):2327-2335. doi:10.1128/jvi.67.4.2327-2335.1993

190. Nestorowicz A, Chambers TJ, Rice CM. Mutagenesis of the Yellow Fever Virus NS2A/2B Cleavage Site: Effects on Proteolytic Processing, Viral Replication, and Evidence for Alternative Processing of the NS2A Protein. Virology. 1994;199(1):114-123. doi:10.1006/viro.1994.1103

191. Chambers TJ, Nestorowicz A, Rice CM. Mutagenesis of the yellow fever virus NS2B/3 cleavage site: determinants of cleavage site specificity and effects on polyprotein processing and viral replication. J Virol. 1995;69(3):1600-1605. doi:10.1128/jvi.69.3.1600-1605.1995

192. Kim JL, Morgenstern KA, Lin C, et al. Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide. Cell. 1996;87(2):343-355. doi:10.1016/S0092-8674(00)81351-3

193. Noble CG, Seh CC, Chao AT, Shi PY. Ligand-Bound Structures of the Dengue Virus Protease Reveal the Active Conformation. J Virol. 2012;86(1):438-446. doi:10.1128/JVI.06225-11

194. Chappell KJ, Stoermer MJ, Fairlie DP, Young PR. Mutagenesis of the West Nile virus NS2B cofactor domain reveals two regions essential for protease activity. J Gen Virol. 2008;89(4):1010-1014. doi:10.1099/vir.0.83447-0

195. de la Cruz L, Chen W-N, Graham B, Otting G. Binding mode of the activitymodulating C-terminal segment of NS2B to NS3 in the dengue virus NS2B-NS3 protease. FEBS J. 2014;281(6):1517-1533. doi:10.1111/febs.12729

196. Erbel P, Schiering N, D'Arcy A, et al. Structural basis for the activation of flaviviral NS3 proteases from dengue and West Nile virus. Nat Struct Mol Biol. 2006;13(4):372-373. doi:10.1038/nsmb1073

197. Kronenberger T, Sá Magalhães Serafim M, Kumar Tonduru A, Gonçalves Maltarollo V, Poso A. Ligand Accessibility Insights to the Dengue virus NS3NS2B Protease Assessed by long-timescale Molecular Dynamics Simulations.
ChemMedChem. May 2021:cmdc.202100246. doi:10.1002/cmdc.202100246

198. Chen WN, Loscha K V, Nitsche C, Graham B, Otting G. The dengue virus NS2B-NS3 protease retains the closed conformation in the complex with BPTI.

FEBS Lett. 2014;588(14):2206-2211. doi:10.1016/j.febslet.2014.05.018

199. Lamiable A, Thévenet P, Rey J, Vavrusa M, Derreumaux P, Tufféry P. PEPFOLD3: faster de novo structure prediction for linear peptides in solution and in complex. Nucleic Acids Res. 2016;44(W1):W449--W454. doi:10.1093/nar/gkw329

200. Santos KB, Guedes IA, Karl ALM, Dardenne LE. Highly Flexible Ligand Docking: Benchmarking of the DockThor Program on the LEADS-PEP Protein–Peptide Data Set. J Chem Inf Model. 2020;60(2):667-683. doi:10.1021/acs.jcim.9b00905

201. Martínez-Rosell G, Giorgino T, De Fabritiis G. PlayMolecule ProteinPrepare: A Web Application for Protein Preparation for Molecular Dynamics Simulations. J Chem Inf Model. 2017;57(7):1511-1516. doi:10.1021/acs.jcim.7b00190

202. Bera AK, Kuhn RJ, Smith JL. Functional Characterization of cis and trans Activity of the Flavivirus NS2B-NS3 Protease. J Biol Chem. 2007;282(17):12883-12892. doi:10.1074/jbc.M611318200

203. Volini M, Tobias P. Circular dichroism studies of chymotrypsin and its derivatives. Correlation of changes in dichroic bands with deacylation. J Biol Chem. 1969;244(19):5105-5109. http://www.ncbi.nlm.nih.gov/pubmed/5344124.

204. Lee WHK, Liu W, Fan JS, Yang D. Dengue virus protease activity modulated by dynamics of protease cofactor. Biophys J. 2021;120(12):2444-2453. doi:10.1016/j.bpj.2021.04.015

205. Xie X, Zou J, Puttikhunt C, Yuan Z, Shi P-Y. Two Distinct Sets of NS2A Molecules Are Responsible for Dengue Virus RNA Synthesis and Virion Assembly. Beemon KL, ed. J Virol. 2015;89(2):1298-1313. doi:10.1128/JVI.02882-14

206. Klaitong P, Smith DR. Roles of Non-Structural Protein 4A in Flavivirus Infection. Viruses. 2021;13(10):2077. doi:10.3390/v13102077

207. Shah PS, Link N, Jang GM, et al. Comparative Flavivirus-Host Protein Interaction Mapping Reveals Mechanisms of Dengue and Zika Virus Pathogenesis. Cell. 2018;175(7):1931--1945.e18. doi:10.1016/j.cell.2018.11.028

208. Ngo AM, Shurtleff MJ, Popova KD, Kulsuptrakul J, Weissman JS, Puschnik AS. The ER membrane protein complex is required to ensure correct topology and stable expression of flavivirus Polyproteins. Elife. 2019;8:1-23. doi:10.7554/eLife.48469

209. Aktepe TE, Liebscher S, Prier JE, Simmons CP, Mackenzie JM. The Host Protein Reticulon 3.1A Is Utilized by Flaviviruses to Facilitate Membrane Remodelling. Cell Rep. 2017;21(6):1639-1654. doi:10.1016/j.celrep.2017.10.055

210. Teo CSH, Chu JJH. Cellular Vimentin Regulates Construction of Dengue Virus Replication Complexes through Interaction with NS4A Protein. J Virol. 2014;88(4):1897-1913. doi:10.1128/JVI.01249-13

211. Harsh S, Fu Y, Kenney E, Han Z, Eleftherianos I. Zika virus non-structural protein NS4A restricts eye growth in Drosophila through regulation of JAK/STAT signaling. Dis Model Mech. January 2020. doi:10.1242/dmm.040816

212. Hu Y, Dong X, He Z, et al. Zika virus antagonizes interferon response in patients and disrupts RIG-I–MAVS interaction through its CARD-TM domains. Cell Biosci. 2019;9(1):46. doi:10.1186/s13578-019-0308-9

213. Lindenbach BD, Rice CM. Genetic Interaction of Flavivirus Nonstructural Proteins NS1 and NS4A as a Determinant of Replicase Function. 1999;73(6):4611-4621.

214. Morita E, Suzuki Y. Membrane-Associated Flavivirus Replication Complex—Its Organization and Regulation. Viruses. 2021;13(6):1060. doi:10.3390/v13061060

215. Yu L, Takeda K, Markoff L. Protein–protein interactions among West Nile nonstructural proteins and transmembrane complex formation in mammalian cells. Virology. 2013;446(1-2):365-377. doi:10.1016/j.virol.2013.08.006

216. Cortese M, Mulder K, Chatel-Chaix L, et al. Determinants in Nonstructural Protein 4A of Dengue Virus Required for RNA Replication and Replication Organelle Biogenesis. López S, ed. J Virol. 2021;95(21). doi:10.1128/JVI.0131021

217. Kaufusi PH, Kelley JF, Yanagihara R, Nerurkar VR. Induction of endoplasmic reticulum-derived replication-competent membrane structures by West Nile virus non-structural protein 4B. PLoS One. 2014;9(1). doi:10.1371/journal.pone.0084040

218. Stern O, Hung Y-F, Valdau O, et al. An N-Terminal Amphipathic Helix in Dengue Virus Nonstructural Protein 4A Mediates Oligomerization and Is Essential for Replication. J Virol. 2013;87(7):4080-4085. doi:10.1128/JVI.01900-12

219. Karimova G, Pidoux J, Ullmann A, Ladant D. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. Proc Natl Acad Sci U S A. 1998;95(10):5752-5756. doi:10.1073/pnas.95.10.5752

220. Ujwal R, Abramson J. High-throughput Crystallization of Membrane Proteins Using the Lipidic Bicelle Method. J Vis Exp. 2012;(59). doi:10.3791/3383

221. Arias-Arias JL, MacPherson DJ, Hill ME, Hardy JA, Mora-Rodríguez R. A fluorescence-activatable reporter of flavivirus NS2B–NS3 protease activity enables live imaging of infection in single cells and viral plaques. J Biol Chem. 2020;295(8):2212-2226. doi:10.1074/JBC.RA119.011319

222. Li Y, Lee MY, Loh YR, Kang C. Secondary structure and membrane topology of dengue virus NS4A protein in micelles. Biochim Biophys Acta - Biomembr. 2018;1860(2):442-450. doi:10.1016/j.bbamem.2017.10.016

223. Dubrau D, Tortorici MA, Rey FA, Tautz N. A positive-strand RNA virus uses alternative protein-protein interactions within a viral protease/cofactor complex to switch between RNA replication and virion morphogenesis. Cherry S, ed. PLOS Pathog. 2017;13(2):e1006134. doi:10.1371/journal.ppat.1006134

224. Yost SA, Marcotrigiano J. Viral precursor polyproteins: keys of regulation from replication to maturation. Curr Opin Virol. 2013;3(2):137-142. doi:10.1016/j.coviro.2013.03.009

225. Lee LT. Structural and interaction studies of the NS4B integral membrane protein from dengue virus. 2017. doi:10.32657/10356/72869