The NS3 protein from dengue virus : structure and function

Luo, Dahai

2009

Luo, D. (2009). NS3 protein from dengue virus: structure and function. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/47439>

<https://doi.org/10.32657/10356/47439>

Nanyang Technological University

Downloaded on 26 Apr 2023 18:39:30 SGT

The NS3 Protein from Dengue Virus: Structure and Function

LUO DAHAI

SCHOOL OF BIOLOGICAL SCIENCES

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy

2009

Acknowledgements

First and foremost, I would like to extend my deepest gratitude to my supervisor, Professor Julien Lescar, for his guidance, encouragement and trust throughout my graduate education.

I am also deeply grateful to Professor Subhash Vasudevan for his supervision and encouragement throughout my graduate education.

My acknowledgements also extend to my collaborators, Professor Gerhard Grüber, Professor Susana Geifman Shochat, Professor Konstantin Pervushin, Professor Jie Yan, Dr Cornelia Hunke, Dr Yin-Hoe Yau, Dr Danny Doan, Mr Leo Wong, Ms Na Wei, Ms Jie Lin, and Mr Wenbo Fu for their excellent work. I must say I truly enjoy working with them.

I also thank Dr Ting Xu, Dr Masayo Kotaka, Dr Siew-Pheng Lim, Dr Thai-Leong Yap, Dr Aruna Sampath, Mr Da-Ying Wen, Mr Wei Liu, Mr Alex Chao, Dr Min Mo, Dr Hui Fan, Dr Amy Ooi, Ms Ravikumar Mano, Ms Hooi-Chen Lee and Mr Bin Wu for their generous help and valuable advices throughout the time I worked with them.

My sincere gratitude also goes to my lab mates, Dr Insaf Qureshi, Dr Shi-Qi Sun, Dr Sujit Dutta, Miss Samantha Yeong, Ms Chiew-Ling Lim, Miss Yee-Hwa Wong, Miss Mitchell Lim, Miss Qi-Wei Huang, Mr Chong-Wai Liew, Mr Joe Yap, Mr Jing-Xian Liu, and Mr Ming-Wei Chen. Without their constant support and help, it would not have been possible for the project to be such a fulfilling and enriching experience for me.

I would like to thank Mr Wong-Boon-Yaw Lee from QES, Mr Kum-Hoong Yip and his colleagues from Bruker for their services on the X-ray machines in SBS NTU, Dr Chern-Hoe Wang from Biopolis for sharing the facilities over there.

Lastly, I am deeply indebted to my family for their love, understanding and support.

Table of Contents

ACKNOWLEDGEMENTS.....	3
TABLE OF CONTENTS.....	4
LIST OF FIGURES.....	6
LIST OF TABLES.....	8
ABBREVIATIONS	9
ABSTRACT	11
CHAPTER 1 INTRODUCTION.....	13
1.1 DENGUE: THE DISEASE AND THE VIRUS.....	13
1.1.1 <i>Dengue, an Emerging Arboviral Disease</i>	13
1.1.2 <i>Infection with dengue viruses</i>	14
1.1.3 <i>Phylogeny of dengue virus</i>	15
1.1.3.1 <i>Flaviviridae</i>	15
1.1.3.2 <i>Flaviviruses</i>	16
1.1.3.3 <i>Dengue virus</i>	17
1.2 STRUCTURAL BIOLOGY OF DENGUE VIRUS.....	18
1.2.1 <i>The Flavivirus life cycle: a structural perspective</i>	18
1.2.1.1 <i>Virus entry</i>	19
1.2.1.2 <i>Fusion and Uncoating</i>	19
1.2.1.3 <i>Translation and polyprotein processing</i>	21
1.2.1.4 <i>Viral genome replication</i>	22
1.2.1.5 <i>Viral particle assembly, maturation and release</i>	23
1.2.2 <i>Structure of dengue virions</i>	24
1.2.3 <i>Structure of the flavivirus genome RNA</i>	27
1.2.4 <i>Structures of the non-structural (NS) proteins</i>	30
1.2.4.1 <i>NS1</i>	30
1.2.4.2 <i>NS2A</i>	31
1.2.4.3 <i>NS2B</i>	32
1.2.4.4 <i>NS3</i>	32
1.2.4.5 <i>NS4A</i>	40
1.2.4.6 <i>NS4B</i>	41
1.2.4.7 <i>NS5</i>	42
1.3 RNA HELICASES: A STRUCTURAL VIEW.....	47
1.3.1 <i>Classification of RNA helicases</i>	48
1.3.2 <i>Conserved motifs</i>	51
1.3.2.1 <i>Motifs involved in ATP binding and hydrolysis</i>	51
1.3.2.2 <i>Motifs involved in RNA binding/remodeling</i>	54
1.3.2.3 <i>Motifs linking ATPase and helicase activities</i>	57
1.3.3 <i>Structure and mechanism of the RNA helicases</i>	58
1.3.3.1 <i>Overall features</i>	58
1.3.3.2 <i>Mechanism of RNA Helicases</i>	63
1.4. AIMS OF THE PROJECT.....	68
1.5 PRESENTATION OF THE WORK.....	69
CHAPTER 2 MATERIALS AND METHODS.....	70
2.1 MOLECULAR CLONING.....	70
2.2 PROTEIN EXPRESSION AND PURIFICATION.....	71
2.3 MATRIX-ASSISTED LASER DESORPTION IONIZATION TIME-OF-FLIGHT MASS SPECTROMETRY.....	73
2.4 N-TERMINAL AMINO-ACID SEQUENCE DETERMINATION.....	73
2.5 CRYSTALLIZATION AND DATA COLLECTION.....	74

Table of Contents

2.6 STRUCTURE SOLUTION AND REFINEMENT	78
2.6 X-RAY SCATTERING EXPERIMENTS AND DATA ANALYSIS OF scNS2B ₁₈ NS3	79
2.8 FLUORESCENCE CORRELATION SPECTROSCOPY EXPERIMENTS	80
CHAPTER 3 RESULTS AND DISCUSSION	84
3.1 CRYSTAL STRUCTURES OF FULL LENGTH NS3 PROTEIN FROM DENV4	84
3.1.1 Expression of scNS2B ₁₈ NS3 and sensitivity to proteolysis	84
3.1.2 Overall structure of scNS2B ₁₈ NS3: Conformation I	86
3.1.2.1 The protease domain	89
3.1.2.2 The helicase domain	91
3.1.3 scNS2B ₁₈ NS3: Conformation II	92
3.1.4 Nucleotide binding properties of scNS2B ₁₈ NS3 and of its isolated helicase domain	95
3.1.5 Small angle X-ray scattering experiments of scNS2B ₁₈ NS3 in solution	97
3.1.6 On the relative orientation between the helicase and protease domains	99
3.2 CRYSTAL STRUCTURES OF DENV4 NS3 HELICASE WITH ITS LIGANDS	103
3.2.1 Structure determination	103
3.2.2 Comparison of DENV4 NS3hel with other flavivirus NS3 helicases	103
3.2.3 ssRNA recognition by DENV4 NS3hel	105
3.2.4 Conformational changes upon ssRNA binding	108
3.2.5 The AMPPNP/Mn ²⁺ and ADP/Mn ²⁺ binary complexes	111
3.2.6 The ATP hydrolysis cycle	112
3.2.7 Coordination of the divalent metal ion	116
3.2.8 An evolutionary conserved ssRNA recognition module	117
3.2.9 Conformational changes during ATP hydrolysis and RNA unwinding	120
3.2.10 Insights into the RNA unwinding mechanism by NS3	121
3.3 INSIGHTS INTO THE VIRAL LIFE CYCLE: ROLES THAT NS3 PLAYS	124
3.3.1 Membrane association via NS2B	124
3.3.2 Implications for polyprotein processing and formation of the replication complex	126
3.3.2.1 NS2B-NS3 protease and viral polyprotein processing	127
3.3.2.2 As part of the viral genome replication complex	129
3.3.3 Non-enzymatic roles played by NS3	134
CHAPTER 4 CONCLUSION AND FUTURE DIRECTIONS	136
REFERENCES	139
APPENDIX	163
A: SOLUTIONS	163
B: PROTEOLYSIS OF scNS2B ₁₈ NS3	164
C: SEQUENCE ALIGNMENT OF THE FLAVIVIRUS NS3 PROTEIN	165
D: CRYSTALLOGRAPHIC CONTACTS IN COMMON BETWEEN scNS2B ₁₈ NS3 CONFORMATION I AND II	167
E: CV_ DAHAI LUO	168

List of Figures

Figure 1. 1 Distribution of dengue viruses and their mosquito vector, <i>Aedes aegypti</i> .	14
Figure 1. 2 Phylogenetic tree of the flaviviruses.	16
Figure 1. 3 A schematic depiction of the Flavivirus life cycle.	18
Figure 1. 4 DENV E protein and class II membrane fusion.	20
Figure 1. 5 Membrane topology of the polyprotein.	22
Figure 1. 6 A model of the flavivirus maturation pathway.	24
Figure 1. 7 Structure of dengue virus particle.	26
Figure 1. 8 Predicted secondary structure of terminal sequences of DENV2 genome.	28
Figure 1. 9 Comparison of subsite specificities of NS3 proteases with their natural cleavage sites in viral polypeptide.	33
Figure 1. 10 Structures of DENV2 NS2B/NS3pro and WNV NS2B/NS3pro with an inhibitor.	35
Figure 1. 11 Structure of DENV2 NS3 helicase.	38
Figure 1. 12 Model of dengue full length NS3.	40
Figure 1. 13 Model for the membrane topology of DENV NS4A.	41
Figure 1. 14 Model for the membrane topology of DENV 2K-NS4B.	42
Figure 1. 15 Schematic representation of the distribution of the functional domains in the DENV NS5 protein.	43
Figure 1. 16 Structure of the flavivirus MTase.	44
Figure 1. 17 Structure representation of DENV 3 RdRp.	47
Figure 1. 18 Sequence and structure organization of RNA helicases.	50
Figure 1. 19 Conserved motifs in the DEAD box and related DExD/H box helicase families.	51
Figure 1. 20 Topology diagrams of RecA and RecA-like core of the SF1 and SF2 helicases.	58
Figure 1. 21 Sketch of helicase on NA.	64
Figure 1. 22 Active rolling model for DNA unwinding and translocation.	64
Figure 1. 23 Inchworm model.	65
Figure 1. 24 Brownian motor mechanism of translocation and unwinding.	66
Figure 1. 25 Three-step mechanism for duplex unwinding by DEAD box helicases.	67
Figure 2. 1 scNS2B ₁₈ NS3 crystals.	75
Figure 2. 2 DENV4 NS3hel crystals.	77
Figure 3. 1 Expression, purification and characterization of scNS2B ₁₈ NS3.	86
Figure 3. 2 Conformation I of the scNS2B ₁₈ NS3 protein from DENV4.	88
Figure 3. 3 Comparison of flaviviridae protease domains.	90
Figure 3. 4 The helicase domain.	92
Figure 3. 5 The relative orientation between the protease and helicase domain of scNS2B ₁₈ NS3.	94
Figure 3. 6 Fluorescence correlation spectroscopy analysis of the interaction between NS3 and the fluorescent nucleotide-analogues.	97
Figure 3. 7 Agreement between the envelope derived from SAXS and the X-ray structure for scNS2B ₁₈ NS3.	99

List of Figures

Figure 3. 8 Interdomain linker and relative orientation between the protease and the helicase domains.	102
Figure 3. 9 Superposition of Flavivirus NS3 helicases.	104
Figure 3. 10 The flavivirus NS3h-RNA-AMPPNP ternary complex.	108
Figure 3. 11 Structural rearrangements upon RNA binding.	110
Figure 3. 12 The AMPPNP/ Mn^{2+} and ADP/ Mn^{2+} binary complex.	112
Figure 3. 13 ATP hydrolysis.	116
Figure 3. 14 Coordination of the divalent ion.	117
Figure 3. 15 An evolutionary-conserved ssRNA recognition module of the DE-A/x-D/H helicases.	119
Figure 3. 16 RNA unwinding by the NS3 Helicase.	122
Figure 3. 17 Model for the membrane bound NS2B-NS3 protein complex.	125
Figure 3. 18 Early events during viral infection.	127
Figure 3. 19 Flavivirus genome 5' cap formation and NS3 RTPase modeling.	130
Figure 3. 20 Model for DENV minus strand RNA synthesis.	133

List of Tables

Table 1. 1 RNA structure elements in flavivirus life cycle..... 29

Table 1. 2 Structures of RNA helicases from SF1/SF2 62

Table 2. 2 Refinement statistics for scNS2B₁₈NS3. 81

Table 2. 3 Data collection Statistics for NS3hel 82

Table 2. 4 Refinement statistics for NS3hel 83

Table 3. 1 Relative orientation of the protease domain. 93

Table 3. 2 Crystallographic contacts in scNS2B₁₈NS3 Conformation I & II. 95

Table 3. 3 Nucleotide binding affinity for scNS2B₁₈NS3 and NS3hel..... 97

Abbreviations

Abbreviation	Full Name
aa	Amino acid
ADP	Adenosine diphosphate
AFM	Atomic force microscopy
ATP	Adenosine triphosphate
ATPase	Adenosine triphosphatase
AMPPNP(ANP)	5'-adenyl- β , γ -imidodiphosphate
β -ME	β -mercaptoethanol
BVDV	Bovine viral diarrhea virus
CCP4	The Collaborative Computational Project Number 4
CDC	Centers for disease control and prevention (United States)
CM	convoluted membrane
CS	Complementary sequence
cryoEM	Cryo-electron microscopy
ds	Double stranded
DF	Dengue fever
DHF	Dengue hemorrhagic fever
DENV	Dengue virus
DTT	Dithiothreitol
DSS	Dengue shock syndrome
EDTA	Ethylenediaminetetra acetic acid
ER	Endoplasmic reticulum
FCS	fluorescence correlation spectroscopy
FL	Full length
HCV	Hepatitis C virus
IPTG	Isopropyl- β -D-thiogalactopyranoside
JEV	Japanese Encephalitis virus
K_d	Dissociation constant
KUNV	Kunjin virus
MALDI-TOF	Matrix-assisted laser desorption ionization time-of-flight
MES	2-morpholin-4-ylethanesulfonic acid
Min	Minute
Mg	Magnesium
Mn	Manganese
MTase	Methyl transferase
MVEV	Murray Valley encephalitis virus
MW	Molecular weight
NA	Nucleic acid
NITD	Novartis Institute for Tropical Diseases (Singapore)
NLS	Nuclear localization signal
NS2B ₄₇	Residues 49-96 of NS2B protein

Abbreviations

NS2B ₁₈	Residues 49-66 of NS2B protein
NS3	Nonstructural protein 3
NS3hel	Nonstructural protein 3 helicase domain
NS3pro	Nonstructural protein 3 protease domain
NTP	Nucleoside triphosphate
nt	Nucleotide
PEG	Polyethylene glycol
PC	Paracrystalline structure
PCR	Polymerase chain reaction
PDB	Protein Data Bank
Pi	Inorganic phosphate
RC	Replication complex
RdRp	RNA dependent RNA polymerase
RF	Replication form
r.m.s. deviation	Root mean square deviation
RT	Room temperature
RTPase	RNA triphosphatase
SAM	S-adenosyl-methionine
SAXS	Small angle x-ray scattering
sc	Single chain
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SL	Stem loop
SF	Superfamily
ss	Single stranded
TBEV	Tick-borne encephalitis virus
TGN	Trans Golgi network
Trx	Thioredoxin reductase
UTR	Untranslated regions
VP	Vesicle packets
WHO	World health organization
WNV	West Nile virus
WT	Wild type
YFV	Yellow Fever virus

Abstract

The Dengue virus belongs to the flavivirus genus which includes other important human pathogens such as Yellow fever, Japanese encephalitis and West Nile virus within the *Flaviviridae* family of enveloped viruses. Dengue virus affects several million people annually, with no effective therapy at hand. The Dengue virion contains a single stranded, positive sense (+) RNA genome of approximately 11 kb which is translated into a large polyprotein containing both structural proteins and nonstructural proteins during the infectious life cycle. The non-structural protein 3 (NS3) is a multifunctional protein of 69 kDa. The N terminal domain of NS3 is a serine protease with NS2B as a cofactor essential for viral polyprotein processing. The C terminal domain is a RNA helicase and a nucleoside 5' triphosphatase, involved in viral replication and virus particle assembly. Thus, NS3 plays an important role in viral life cycle and represents a target for the development of specific antiviral inhibitors.

Atomic structures are available for the individual protease and helicase domains of NS3 separately but a molecular view of the full length flavivirus NS3 polypeptide is still lacking. In order to study its overall architecture and the implications on the enzymatic activities of the NS2B-NS3 protease-helicase, a complete NS3 molecule fused with 18 residues of the NS2B cofactor through a flexible linker (scNS2B₁₈NS3) was engineered. The recombinant protein was successfully expressed, purified and crystallized. The first crystal structure of scNS2B₁₈NS3 was determined as apo enzyme to 3.1 Å. The relative orientation between the protease and helicase domains is drastically different compared to the scNS3-NS4A molecule from hepatitis C virus (HCV), which was caught in the act of *cis* cleavage at the NS3-NS4A junction. For dengue NS3, the protease domain sits beneath the ATP binding site, giving the molecule an elongated shape. The domain arrangement found in the crystal structure fits nicely into an envelope determined *ab-initio* using small angle X-ray scattering experiments in solution. A second conformation of scNS2B₁₈NS3 was also discovered, in which the protease domain rotates ~ 160° relative to the helicase domain in contrast to that in Conformation I.

Abstract

The structural dynamics of scNS2B₁₈NS3 and the potential impacts to the enzymatic activities as well as its roles in viral life cycle will be discussed.

Together with the NS5 polymerase, the NS3 helicase plays a pivotal role in flavivirus RNA replication. The dengue virus NS3 helicase was captured at several stages along the catalytic pathway: ligand-free state, bound to single stranded RNA, to an ATP analog, to a Vanadate-trapped ADP complex and to ATP hydrolysis products. RNA recognition appears largely sequence-independent in a way remarkably similar to eukaryotic DEAD box proteins Vasa and eIF4AIII. An extensive structural rearrangement of NS3 helicase upon ssRNA binding was observed. The NS3 helicase switches to a catalytic-competent state imparted by an inward movement of the P-loop, interdomain closure and a change in the divalent metal coordination shell. The structural basis for RNA-stimulated ATP hydrolysis is suggested based on these observations. In summary, these structures demonstrate for the first time large quaternary changes in the *flaviviridae* helicase, identify the catalytic water molecule for ATP hydrolysis and point to a β -hairpin that protrudes from subdomain 2 as a critical element for dsRNA unwinding. They also suggest how NS3 could act as an RNA anchoring device and thus participate both in flavivirus RNA replication and assembly.

Chapter 1 Introduction

1.1 Dengue: the disease and the virus

1.1.1 Dengue, an Emerging Arboviral Disease

The diseases caused by dengue virus (DENV) infection are dengue fever (DF) and its more serious forms, dengue hemorrhagic fever (DHF) and dengue shock syndrome (DSS) which have become a major international public health concern in recent decades (Gibbons and Vaughn, 2002; Gubler, 1998). Dengue is found in tropical and sub-tropical regions around the world, predominantly in urban and semi-urban areas. The DENV are transmitted among humans by infective female mosquitoes of the genus *Aedes*. The most efficient dengue epidemic vector is *Aedes aegypti* (Figure 1. 1). Four distinct but antigenically related serotypes (DENV1-4) of dengue are documented since the first report of isolation of DENV1 by Japanese and American investigators in the Pacific during World War II (Calisher *et al.*, 1989; Hotta, 1952a, b; Sabin, 1952).

The incidence of dengue has grown dramatically around the world in recent decades (**Figure 1. 1**). Some 2.5 billion people – two fifths of the world's population – are now at risk from dengue. World Health Organization (WHO) currently estimates that there may be 50 million dengue infections worldwide every year. In 2007 alone, there were more than 890 000 reported cases of dengue in the Americas, of which 26 000 cases were DHF. The disease is now endemic in more than 100 countries all around the world. South-east Asia and the Western Pacific are the most seriously affected (WHO, 2009). Before 1970, only nine countries experienced DHF epidemics. The number is increased more than four-fold by 1995. In 2007, Venezuela reported over 80 000 cases, including more than 6 000 cases of DHF (WHO, 2009).

The spread of dengue is attributed to expanding geographic distribution of the four dengue viruses and their vectors; the most important of which is the urban species *Aedes aegypti*. A rapid rise in urban mosquito populations is bringing ever greater numbers of people into contact with this vector, especially in areas that are favorable for mosquito breeding, where household water storage is common and where solid waste disposal services are inadequate (WHO, 2009).

There is no specific treatment for dengue currently available, although appropriate medical care frequently saves the lives of patients with DHF. The only way to prevent dengue virus transmission is to combat the vector carrying the disease (WHO, 2009).

Figure 1. 1 Distribution of dengue viruses and their mosquito vector, *Aedes aegypti*. Left panel: Geographical distribution of dengue (light shading) and DF / DHF (dark shading) (2005); Right panel: *Aedes aegypti* taking blood meal. Adapted from CDC website.

1.1.2 Infection with dengue viruses.

Dengue virus infection can lead to a range of symptoms, from mild DF to severe DHF and DSS. Dengue fever is a flu-like illness that affects infants, young children and adults, but seldom causes death. The clinical features of dengue fever vary according to the age of the patient. Infants and young children may have a fever with rash. Older children and adults may have either a mild fever or the classical incapacitating disease with abrupt onset and high fever, severe headache, pain behind the eyes, muscle and joint pains, and rash (Halstead, 1988; Halstead *et al.*, 2005; Henchal and Putnak, 1990). DHF is a potentially deadly complication that is characterized by high fever, often with enlargement of the liver and in severe cases circulatory failure. The illness often begins with a sudden rise in temperature accompanied by facial flush and other flu-like symptoms. The fever usually continues for two to seven days and can be as high as 41°C, possibly with convulsions and other

complications. In severe cases, the patient's condition may suddenly deteriorate after a few days of fever; the temperature drops, followed by signs of circulatory failure and the patient may rapidly go into a critical state of shock and die within 12 to 24 hours or quickly recover following appropriate medical treatment to maintain the patient's circulating fluid volume (Halstead, 1988; Halstead *et al.*, 2005; Henchal and Putnak, 1990).

Although the mechanisms for the development of DHF and DSS are not fully understood, one risk factor for the development of DHF and DSS is thought to be secondary infection with another serotype (Gubler, 1998; Guzman *et al.*, 2002). The cross-reactive but non-neutralizing anti-dengue antibodies from the previous infection bind to the new infecting serotype and enhance viral uptake of monocytes and macrophages, via antibody-dependent enhancement (ADE) (Halstead, 1979; Halstead *et al.*, 1983). ADE results in an amplified cascade of immune responses involving cytokines and complement activation, causing platelet destruction, endothelial dysfunction, and consumption of coagulation factors leading to plasma leakage and hemorrhagic manifestations (Guzman and Kouri, 2002; Halstead, 1979, 1983; Halstead *et al.*, 1983).

1.1.3 Phylogeny of dengue virus

1.1.3.1 Flaviviridae

Flaviviridae is a family of (+) RNA viruses that are responsible for a broad spectrum of pathogenic manifestations in humans, domestic animals and birds (Heinz *et al.*, 2000). It consists of three genera: hepacivirus (e.g, hepatitis C virus, HCV), pestivirus (e.g, bovine viral diarrhea virus, BVDV) and flavivirus (e.g, Dengue virus, DENV). In addition, a group of unassigned viruses, the GB agents, are to be formally classified within the family. The genome of the Flaviviridae viruses is a linear, single-stranded RNA of about 10,000-11,000 nucleotides long. Both 5' and 3' ends of the genome are untranslated regions (UTR). The viral particle is enveloped and spherical, about 40-60 nm in diameter. Although they share similarities in virion morphology, genome organization and RNA replication strategy, members within this family exhibit diverse biological properties and lack of serologic cross-reactivity. The increasing significance of Flaviviridae as human and animal pathogens makes it an

increasing interest to the academic community (Heinz *et al.*, 2000; Lindenbach *et al.*, 2007).

1.1.3.2 Flaviviruses

The name flaviviruses comes from the Latin word “flavus,” meaning yellow that signifies jaundice, a common sign of infection with the prototypic yellow fever virus, which is among the first filterable agents shown to cause human diseases (along with DENV) (Ashburn and Craig, 2004; Reed and Carroll, 2001).

Figure 1. 2 Phylogenetic tree of the flaviviruses. The phylogenetic tree is derived from partial NS5 sequences from the GenBank database. Subtypes are written in parentheses after virus names. New World viruses are printed in bold and underlined. Numbers indicate bootstrap values for major clades to the right. Adapted from Vasilakis and Weaver, 2008.

Viruses in the Flavivirus genus are grouped taxonomically into three groups according to their vector association and antigenic relationships: (1) tick-borne (2) mosquito borne and (3) viruses with no known arthropod vector (Vasilakis and Weaver, 2008). They are widely distributed throughout the world. Several members of the Flavivirus genus, e.g., dengue virus (DENV), yellow fever virus (YFV) Japanese encephalitis virus (JEV), tick-borne encephalitis virus (TBEV) and West Nile virus (WNV) are well known medically important pathogens that affect humans (Figure 1. 2).

1.1.3.3 Dengue virus

DENV is considered a species within the genus Flavivirus and includes four distinct but antigenically related serotypes. Nucleic acid sequencing allows the classification of DENV into genetically distinct groups or genotypes within each serotype (Rico-Hesse, 1990). Various phylogenetic analyses based on partial E/NS1, complete E nucleotide sequences or complete genome sequences indicate that DENV1 is grouped in five genotypes (1) genotype I represents strains from Southeast Asia, China, and East Africa, (2) genotype II represents strains from Thailand collected in the 1950s and 1960s, (3) genotype III represents the sylvatic strain collected in Malaysia (4) genotype IV represents strains from the West Pacific islands and Australia and (5) genotype V represents all strains collected in the Americas, strains from West Africa, and a limited number of strains collected from Asia (Goncalvez *et al.*, 2002; Rico-Hesse, 1990). Similarly, DENV2 is grouped in five genotypes, DENV3 five genotypes and DENV4 four genotypes (Vasilakis and Weaver, 2008).

The evolutionary relationship of DENV remains elusive, although evidence supports the hypothesis that all four DENV serotypes emerged into the human population independently from sylvatic progenitors that circulate in forests of Asia among nonhuman primates via arboreal mosquito vectors (Gubler *et al.*, 1979; Holmes and Twiddy, 2003; Rico-Hesse, 1990; Wang *et al.*, 2000).

1.2 Structural biology of dengue virus

Due to the advances on the various techniques and the research efforts, the structure biology of flavivirus has entered a new era. Our knowledge about the viral proteins, viral RNA, viral particle and the virus life cycle is rapidly growing. For the interest of this writing, dengue virus and other closely related flaviviruses are the primary focus of this section.

1.2.1 The Flavivirus life cycle: a structural perspective

Figure 1. 3 A schematic depiction of the Flavivirus life cycle. (1) Virus entry by receptor mediated endocytosis. (2) Membrane fusion induced by acidic pH in the endosome and release of the nucleocapsid. (3) Uncoating. (4) Translation of the plus stranded RNA genome to initiate replication, co-/post- translational proteolytic processing of viral polyprotein by viral and host proteases, and replication of the virus genome. (5) Virus assembly in the ER and transportation of immature virion to the Golgi apparatus. (6) Proteolytic cleavage of prM in trans Golgi network (TGN) and maturation of the virus. (7) Release of the infectious virus particles and subviral particles (**SHA**, slowly-sedimenting hemagglutinin) by exocytosis. Adapted from Stiasny and Heinz, 2006.

1.2.1.1 Virus entry.

Flavivirus particle infects host cells via receptor-mediated endocytosis that is triggered by the interaction between the viral surface glycoproteins and cellular receptors. Depending on the target cells, the primary receptors have been reported to be dendritic-cell-specific ICAM-grabbing non-integrin (DC-SIGN) (Navarro-Sanchez *et al.*, 2003), glucose-regulating protein 78 (GRP78/Bip) (Jindadamrongwech *et al.*, 2004) and lipopolysaccharide CD14 associated molecules (Chen *et al.*, 1999). In addition, glycosaminoglycans (GAGS) and heparin sulfate are low affinity co-receptors for several flaviviruses (Chen *et al.*, 1997b; Germe *et al.*, 2002; Kroschewski *et al.*, 2003; Lee and Lobigs, 2002). Unlike many other viruses, very little structural information about the mechanisms by which flaviviruses bind to their receptors is available (Dimitrov, 2004).

1.2.1.2 Fusion and Uncoating.

Membrane fusion occurs within the endosome, mediated by viral surface E protein. Upon acidification, the E protein dimers undergo extensive conformational changes to form trimers (Allison *et al.*, 1995; Rey *et al.*, 1995; Stiasny *et al.*, 2001; Stiasny *et al.*, 1996); the fusion loops are then exposed and inserted into the endosomal membrane; the endosomal and viral membranes are brought into close proximity to fuse due to further structural rearrangements of E trimers (Modis *et al.*, 2004). This is known as class II membrane fusion process (Reviews on class I and class II membrane fusion, (Mukhopadhyay *et al.*, 2005; Stiasny and Heinz, 2006). NC is then disassembled probably due to acidification and the viral genome is then released into the host cell cytoplasm.

Figure 1. 4 DENV E protein and class II membrane fusion. (A) Dimeric, pre-fusion conformation of the DENV2 E protein. In one monomer, domains I, II and III are in red, yellow and blue, respectively. The fusion peptide is shown in green. The other monomer is in grey. (B) Trimeric, post-fusion conformation of the DENV2 E protein. One monomer has the individual domains and fusion peptide coloured as in part A, whereas the other two monomers are in grey. (C) Class II fusion. The dimeric E protein is on the virus surface with the fusion peptide (shown in green) buried in the dimer (a). The protein binds to a receptor and is internalized into the endosome. Under low pH, domain II swings outward towards the host cell membrane, presumably at the domain I–II hinge region (b). This conformational change allows the E proteins to rearrange laterally. The fusion peptide inserts itself into the outer leaflet of the host cell membrane, enabling trimerization of the E protein (c). Domain III of the E protein folds back onto itself, and in the process brings the viral membrane towards the fusion peptide and host cell membrane (d). As domain III moves towards domain II, hemifusion of the lipid membranes occurs (e), and finally a trimer forms where the transmembrane regions and the fusion peptide are in close proximity (f). Adapted from Mukhopadhyay *et al.*, 2005.

1.2.1.3 Translation and polyprotein processing.

Once reaching the cytoplasm of the host cell, (+) viral genomic RNA serves as mRNA immediately for translation of viral polyprotein (Lindenbach *et al.*, 2007). The resulted single chain polyprotein is co- and post-translationally processed by viral and host proteases into 10 viral proteins (**Figure 1. 5**). Convoluted membranes (CMs) and paracrystalline structures (PCs), probably derived from ER, are likely to be the subcellular locations where polyproteins were deposited and processed (Mackenzie and Westaway, 2001; Westaway *et al.*, 2003). The cleavages at N terminal structural proteins (C/prM, prM/E, E/NS1) and the C terminal of NS4A in the ER lumen are mediated by host signal peptidase. Processing of the C terminal non-structural (NS) proteins (NS2A/NS2B, NS2B/3, NS3/NS4A, and NS4B/NS5) at the cytoplasmic side of ER membrane are mediated by the NS2B/NS3 protease (Lindenbach *et al.*, 2007). In addition, internal cleavage in C protein by the NS2B-NS3 protease is a prerequisite for the efficient cleavage at the C/prM junction (Stocks and Lobigs, 1998). Similarly, internal cleavage in NS4A protein by the NS2B-NS3 protease is also prerequisite for the efficient cleavage at the downstream NS4A/NS4B junction (Lin *et al.*, 1993a). There is also an internal cleavage next to the C-terminal of NS2A by NS2B-NS3 protease (Chambers *et al.*, 1990c; Nestorowicz *et al.*, 1994). Inside the *trans* Golgi network (TGN), a cellular protease furin cleaves prM on the surface of immature viral particles to generate the mature virus (Li *et al.*, 2008; Stadler *et al.*, 1997; Yu *et al.*, 2008a).

Figure 1. 5 Membrane topology of the polyprotein. The viral RNA is translated as a polyprotein and processed by cellular and viral proteases (denoted by arrows). The structural proteins include capsid (C), membrane protein (prM/M) and envelope (E). prM and E are released from the polyprotein by signalase cleavage in the ER, but remain anchored on the luminal side of the membrane. The C is also anchored in the ER membrane (on the cytoplasmic side) by a conserved hydrophobic signal sequence at its C-terminal end. This signal sequence is cleaved by the viral NS2B–NS3 protease. During virus maturation, prM is further cleaved by furin in the TGN into the pr peptide and M protein. The non-structural proteins are processed mainly by the viral protease NS2B–NS3 in the cytoplasm with the exception of NS1, which is released from NS2A by an unidentified protease in the lumen of the ER. NS2A/2B and NS4A/4B are anchored in the ER as transmembrane proteins. The topology of NS4A and NS4B are predicted through biochemical and cellular analyses. Adapted from Perera and Kuhn 2008.

1.2.1.4 Viral genome replication.

Once viral proteins are available after initial rounds of polyprotein translation and processing, NS proteins start to assemble into viral replication complex. Viral genome replication is catalyzed within the viral replication complex that comprises (1) NS5, RNA methyl transferase (MTase) at the N terminal region and RNA dependent RNA polymerase (RdRp) at the C terminal region; (2) NS3, the viral RNA helicase at the C terminal; (3) Viral RNA, other NS proteins (i.e. NS1, NS2A, and NS4A) (Ahlquist *et al.*, 2003; Bartholomeusz and Wright, 1993; Chen *et al.*, 1997a; De Nova-Ocampo *et al.*, 2002; Edward and Takegami, 1993; Lindenbach and Rice, 1997, 1999; Mackenzie *et al.*, 1998; Muylaert *et al.*, 1997; Westaway *et al.*, 2003). First, (-) anti-genomic RNA is synthesized resulting in the formation of the double stranded RNA, Replication Form (RF). Next, asymmetric and semi-conservative RNA synthesis then lead to 10 times more (+) strand production than that of (-) strands (Chu and Westaway, 1985, 1987). No subgenomic virus-specific RNAs have been detected in flavivirus-infected cells, in support of the concept that the complete genomic (+) RNA is the only virus-specific mRNA. It is increasingly evident that ER derived membrane structures (smooth membrane vesicle-like structures SMS, also

called vesicle packets, VP) (Mackenzie and Westaway, 2001) are involved in the assembly of the viral replication particle and viral RNA replication, which is likely to be a common replication strategy shared among many different (+) RNA viruses (Denison, 2008; Salonen *et al.*, 2005; Uchil and Satchidanandam, 2003a, b; Westaway *et al.*, 2003).

1.2.1.5 Viral particle assembly, maturation and release.

Immediately after viral protein expression and genome replication, the progeny viral genome RNA associates with capsid proteins forming nucleocapsid. Viral replication and packaging are closely related, supported by encapsidation studies with KUNV replicon RNA (Khromykh *et al.*, 2001b; Khromykh *et al.*, 1998; Varnavski and Khromykh, 1999). Assembly of virions occurs within the lumen of the rough ER. Subsequently, maturation takes place in the secretory pathway (**Figure 1. 6**) (Mackenzie and Westaway, 2001). In TGN, the maturation of virion occurs by processing of prM protein on the immature viral particle surface (Li *et al.*, 2008; Yu *et al.*, 2008a). Mature viruses are then released from the infected cell through exocytosis. Interestingly, apart from their involvement in viral RNA replication, NS proteins are shown to participate in virus morphogenesis (Murray *et al.*, 2008). Mutagenesis studies have shown NS2A has an important role in the incorporation of genomic RNA into the budding virion in YFV and in the production of infectious WNV virus (Kummerer and Rice, 2002; Leung *et al.*, 2008; Liu *et al.*, 2003). Similarly, NS3 has similar functional involvement in virus assembly independent of its known enzymatic activities (Liu *et al.*, 2003; Patkar and Kuhn, 2008).

Figure 1.6 A model of the flavivirus maturation pathway. (A) The conformational changes of the virus particles in the secretory pathway. Immature particles bud into ER as spiky virions and are transported through golgi into the TGN, where acidification induces a conformational change of the virion. Furin cleavage takes place in the TGN, and pr remains associated until the virion is released to the extracellular milieu. The approximate luminal pH values of the specified cellular compartment are indicated. (B) Configuration of the glycoproteins on the surface of the virion during maturation. The structure of the E protein in the secretory pathway is largely unchanged, except for movements at the hinge between domains I and II. In contrast, the oligomerization states of the glycoproteins are critically dependent on pH. The fusion loops are indicated by red stars. Adapted from Yu *et al.*, 2008a.

1.2.2 Structure of dengue virions

Flavivirus virions are approximately 500 Å in diameter and are composed of a single, (+) RNA genome that is packaged together with virus capsid protein (NC, nucleocapsid) in a host-derived lipid bilayer and surrounded by 180 copies each of two glycoproteins (E, envelope protein; M, membrane protein) (Lindenbach *et al.*, 2007).

The structure of dengue virus particle was determined at a resolution of 24Å using cryo-electron microscopy (cryoEM) reconstruction (**Figure 1. 7 A/B**) (Kuhn *et al.*, 2002). Mature dengue virus particles have a smooth surface and icosahedral symmetry (**Figure 1. 7 C**). The outermost, thin, ordered and dense layer of the particle (radii of 220–245 Å) corresponds to the E and M glycoproteins. The next shell down, between radii 185 and 220 Å, is the ectodomains of the M proteins and the stem regions of the E glycoproteins. Beneath the protein shells, the viral membrane bilayer has a radii between 140 and 185 Å. Internal to the lipid region is the nucleocapsid shell situated between radii of 105 and 135 Å, which has a less ordered structure. The maximum height of the density of the nucleocapsid shell is only 50% of that of the external glycoprotein. This suggests that the core protein is poorly ordered or that the orientation of the core as a whole is somewhat variable relative to the external glycoprotein scaffold. By fitting tick-borne encephalitis E protein structure (Rey *et al.*, 1995) into the density, the external surface of the dengue virus particle is revealed to be composed of 90 E dimers arranged in an unusual “herring-bone” manner with three monomers per asymmetric unit (Kuhn *et al.*, 2002).

In contrast to the smooth surface of the mature virus, immature virions in which the prM protein has not been cleaved, have a rough surface characterized by 60 spikes each consisting of 3 prM-E heterodimers that reach to an external diameter of 600 Å. In each prM-E heterodimer, E protein protrudes away from the viral membrane making its long axis tilt 25° angle with respect to the viral surface. The fusion peptide on its outer tip is covered by the N-terminal portion of prM, consistent with the notion that prM protects fusion peptide during maturation. The transformation from immature to mature particles requires some large rearrangement of the E and M proteins.

The structure of mature dengue virus particles was later determined to a resolution of 9.5 Å by improved cryo-EM and image reconstruction techniques. It is then possible to determine the topology of the 180 envelope (E) and 180 membrane (M) proteins in the lipid envelope (**Figure 1. 7 D**) (Zhang *et al.*, 2003).

Figure 1. 7 Structure of dengue virus particle. (A) DENV2 cryoEM reconstruction at 24 Å resolution shown as a surface-shaded representation. Outline of one icosahedral asymmetric unit and the definition of the coordinate system are highlighted. Scale bar represents 100 Å. (B) Cross-section showing the cryoEM density with a plot of the maximum (blue) and averaged (purple) density. Arrow indicates the position of the 5-fold and 3-fold axes. Radial density sections shown at the defined radii, r1: nucleocapsid; r2: lipid bilayer; r3: M glycoprotein; r4: E glycoprotein. Dark shading represents protein in higher density. Scale bar represents 175 Å. (C) Mature dengue virus particle. Three copies of E dimers are highlighted. Adapted from Kuhn *et al.*, 2002. (D) Diagrams of the dengue virus ectodomain and transmembrane domain proteins. The volume occupied by the ectodomain of an E monomer is pink (domain I), yellow (domain II) and blue (domain III). The stem and

anchor helices of E and M are blue and orange, respectively. Adapted from Zhang *et al.*, 2003.

1.2.3 Structure of the flavivirus genome RNA

Flaviviruses contain a (+) strand RNA genome that is approximately 10,500 nucleotides in length. The RNA contains a type 1 cap (m7GpppAmp-) at its 5' terminal (Cleaves and Dubin, 1979) but is not polyadenylated (Wengler, 1981), and encodes a single, long open reading frame (ORF) flanked by 5' and 3' untranslated regions (UTR) (Chambers *et al.*, 1990a). The 5' UTRs are about 100 nucleotides long, while the 3' UTRs vary between 350 and 700 nucleotides. Viral proteins are derived by co- or post-translational cleavage of the polyprotein encoded by the ORF.

Viral RNA genomes not only code for proteins, but in many instances also carry RNA motifs that play a crucial role in the viral life cycle. The first complete flavivirus genome sequence published was that of the YFV (Rice *et al.*, 1985). The extreme 5'- and 3'- terminal sequences of the YFV genome were observed to be partially homologous to those of the WNV (Wengler, 1981). Since then, the structure and function of these essential noncoding sequences in relation to the virus life cycle have been subjected to extensive research (Markoff, 2003). The 5' and 3' UTRs are known to form specific secondary structures and complicated long-range interactions like genome cyclization (**Figure 1. 8**) (Villordo and Gamarnik, 2008). They have been implicated in the initiation of (-) strand synthesis on virus RNA released from entering virions, switching from (-) strand synthesis to synthesis of progeny (+) RNA at late times after infection, and possibly in the initiation of translation and in the packaging of virus plus strand RNA into particles (Markoff, 2003). However, none of these mechanisms is evidently defined.

Conserved sequence features of 5' and 3' NTR in flavivirus RNA have been identified and examined by computational studies on RNA sequence homogeneity and RNA secondary structure folding prediction and by *in vitro* / *in vivo* experiments with engineered RNAs and virus infectious clones (Markoff, 2003; Thurner *et al.*, 2004). Well known conserved structural elements are 5'-AG/UC-3', 5' stem loop A (SLA), 5' stem loop B (SLB), 5' capsid region hairpin (cHP), and the 3' stem loop (3'SL). 5'-3' complementary sequences (CS) and 5'-3' Upstream AUG Region (UAR) are long-

range RNA-RNA contacts mediating viral genomic RNA cyclization and the formation of the “panhandle” RNA structure (**Figure 1. 8**) (Polacek *et al.*, 2009; Villordo and Gamarnik, 2008). The genome cyclization was directly visualized at single molecule level using atomic force microscopy (AFM) (Alvarez *et al.*, 2005b) and was implicated in viral RNA synthesis (Villordo and Gamarnik, 2008). The description and their functional insights of the RNA structure elements in viral life cycle are summarized in **Table 1. 1**. Clearly, these viral RNA sequences / structures play different roles at almost all stages of the virus life cycle. The core functionality is to direct the replication of the viral genome itself, from the formation of the replication complex, initial (-) strand RNA synthesis, formation of RF and generation of (+) strand RNA progeny in large quantity.

Figure 1. 8 Predicted secondary structure of terminal sequences of DENV2 genome. Top panel, structural elements located at the 5' end of the genome: stem loop A (SLA), stem loop B (SLB) and capsid region hairpin (cHP) are shown linked by a dash line, representing the viral genome, to the 3' stem loop (3'SL). Lower panel

shows the changes upon hybridization of the complementary sequences, which results in RNA cyclization. Note that the predicted SLA and cHP are maintained while the SLB and the 3' SL changed after 5'-3' UAR (Upstream AUG Region) and 5'-3' CS interaction. Adapted from Villordo and Gamarnik, 2008.

Table 1. 1 RNA structure elements in flavivirus life cycle

Sequence / Structure	Description	Functional Insights	References
5'-AG/UC-3'	Conserved dinucleotide sequences present at the ends of the genome or anti genome of all vector-borne flavivirus.	<ul style="list-style-type: none"> - The penultimate C at the 3' terminal of KUN RNA (+) and (-) strands was essential for KUN/WNV RNA replication, but not for viral translation. - Mutations that changed the 3'-terminal nucleotide (uridine) of either strand blocked (U to A/G) or restricted replication (U to C). - 3' terminal serve as contact site for specific assembly of the replication complex or for efficient initiation of (-) sense RNA synthesis. 	(Khromykh <i>et al.</i> , 2003; Markoff, 2003; Rice <i>et al.</i> , 1985; Tilgner and Shi, 2004; Wengler, 1981)
5' stem loop A (SLA)	Stable Y shape stem and loop structure consisted of a large stem loop with a side loop; unchanged conformation upon genome cyclization.	<ul style="list-style-type: none"> - Essential for RNA dependent RNA polymerase (RdRp) NS5 recognition as the first step for initiation of (-) strand RNA synthesis and promote the downstream RNA synthesis. - Top loop contains nucleotide sequence specificity for both polymerase activity and viral replication. 	(Filomatori <i>et al.</i> , 2006; Yu <i>et al.</i> , 2008b)
5' stem loop B (SLB)	A short stem loop containing the 5'UAR; Conformation is changed upon genome cyclization.	<ul style="list-style-type: none"> - 5'SLB is not determinant, but dispensable for 5'UTR interaction with NS5. - It might play regulatory roles in switching between translation and RNA synthesis. 	(Filomatori <i>et al.</i> , 2006; Polacek <i>et al.</i> , 2009; Villordo and Gamarnik, 2008)
5' capsid region hairpin (cHP)	A stable stem loop within the capsid coding region, not belonging to the 5'UTR.	<ul style="list-style-type: none"> - 5'cHP directs viral protein translation start site selection in human and mosquito cells. - The ability of the cHP to direct initiation from the first start codon is proportional to its thermodynamic stability, is position dependent, and is sequence independent, consistent with a mechanism in which the scanning initiation complex stalls momentarily over the first AUG as it begins to unwind the cHP. - required for efficient viral RNA synthesis in a sequence-independent manner 	(Clyde <i>et al.</i> , 2008; Clyde and Harris, 2006)
3' stem loop (3'SL)	3'-terminal sequence (approximately 100 nucleotides) form a highly stable stem and loop structure; Conformation is changed upon genome cyclization.	<ul style="list-style-type: none"> - Interacting with NS5 RdRp and NS3 helicase, important for (-) strand RNA synthesis; - Binding to several cellular proteins required for viral replication e.g., the translation elongation factor EF1α to modulate viral protein translation. 	(Blackwell and Brinton, 1997; Chen <i>et al.</i> , 1997a; Markoff <i>et al.</i> , 2002; Villordo and Gamarnik, 2008;

Chapter 1 Introduction

		- Mutations of the 3' SL also affect the host range of the virus.	Yu <i>et al.</i> , 2008b; Zeng <i>et al.</i> , 1998)
5'-3' complementary sequences (CS)	The 5'CS lies amino terminal part of capsid protein; 3'CS lies within 3'UTR, upstream of the 3'SL. Formation of long range RNA-RNA interaction and cyclization of the genome.	- Complementarity between 5' and 3' CS is essential for viral replication, demonstrated using in vitro NS5 polymerase activity assay, infectious clones and replicon systems of KUNV, DENV, and WNV. - Translation on the input RNA is not affected by the alternation of the long range RNA-RNA interaction. - Unknown if the genome cyclization is also involved in later rounds of viral protein translation or viral genome encapsidation.	(Alvarez <i>et al.</i> , 2005a; Alvarez <i>et al.</i> , 2005b; Clyde and Harris, 2006; Hahn <i>et al.</i> , 1987; Khromykh <i>et al.</i> , 2001a; Lo <i>et al.</i> , 2003; Villordo and Gamarnik, 2008; You and Padmanabhan, 1999)
5'-3' Upstream AUG Region (UAR)	5' UAR is upstream of the translation initiator AUG at the 5' UTR, part of the 5'SLB; 3' UAR is present within the stem of the 3' SL. Formation of long range RNA-RNA interaction and cyclization of the genome.	- Complementarity between 5' and 3' UAR is also essential for viral RNA synthesis in DENV and WNV. - It might be playing regulatory roles in switching between translation and RNA synthesis.	(Alvarez <i>et al.</i> , 2008; Zhang <i>et al.</i> , 2008)

1.2.4 Structures of the non-structural (NS) proteins

Seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) are translated and processed from the C terminal of the viral polyprotein (Lindenbach *et al.*, 2007). Non-structural proteins are extensively involved in various stages of virus life cycle: polyprotein processing, formation of viral replication complex, viral RNA synthesis and virus assembly (Lindenbach *et al.*, 2007; Murray *et al.*, 2008). Structures of NS2B-NS3 protease (Erbel *et al.*, 2006), NS3 RNA helicase (Wu *et al.*, 2005; Xu *et al.*, 2005), NS5 methyltransferase (Egloff *et al.*, 2002), and NS5 RdRp (Malet *et al.*, 2007; Yap *et al.*, 2007) were determined and their enzymatic activities were studied. The other NS proteins contain no known enzymatic activities and there is insufficient structural information about them. In addition, the structure of full length NS5 is also yet to be determined.

1.2.4.1 NS1

NS1 (46kDa) is translated into the ER during synthesis and cleaved from E protein by a host signal peptidase, whereas an unknown ER-resident host enzyme

cleaves the NS1/2A junction (Chambers *et al.*, 1990b; Falgout *et al.*, 1989; Falgout and Markoff, 1995). In the ER, the NS1 hydrophilic monomer forms hydrophobic non-covalent homodimer and interacts peripherally with membrane components to be involved in viral replication (Lindenbach and Rice, 1997, 1999; Parrish *et al.*, 1991; Winkler *et al.*, 1989). NS1 contains two or three N-linked glycosylation sites and six disulfide bonds formed by 12 conserved cysteines (Lee *et al.*, 1989; Mason, 1989). N-linked glycosylation of NS1 of dengue virus involves two conserved asparagines, Asn-130 and Asn-207, which are important for NS1 dimer stability and membrane association (Pryor and Wright, 1994; Winkler *et al.*, 1988). Arrangement of the disulfide linkage has been mapped to be C1/C2, C3/C4, C5/C6, C7/C12, C8/C10 and C9/C11 for dengue NS1 protein (Wallis *et al.*, 2004). Alternatively, NS1 exists either as cell-associated via a glycosyl-phosphatidylinositol group (GPI) or secreted form as a soluble hexamer (sNS1) from infected mammalian cells (Falgout *et al.*, 1989; Flamand *et al.*, 1999; Winkler *et al.*, 1989; Winkler *et al.*, 1988). Interestingly, secreted NS1 induces strong immunological responses (Chung *et al.*, 2006; Henchal *et al.*, 1988; Lin *et al.*, 2005). The importance of NS1 to the flavivirus life cycle and immune response makes it an attractive target for development of antiviral therapeutics.

1.2.4.2 NS2A

NS2A (22 kDa) is a small hydrophobic integral membrane protein. Cleavage at the N-terminal NS1/NS2A junction is within the ER lumen whereas the C-terminal is generated by the NS2B/NS3 protease at the cytoplasmic side of the ER (Falgout and Markoff, 1995). An internal cleavage next to the C-terminal of NS2A by the viral NS2B/NS3 protease leads to a truncated form NS2A_α (20kDa) (Chambers *et al.*, 1990b; Nestorowicz *et al.*, 1994). Mutagenesis studies on the internal cleavage site mutants Lys-190 of YFV NS2A indicate it plays an important role in infectious particle production other than NS2A_α cleavage site recognition by viral protease (Kummerer and Rice, 2002). The defect in virus particle production caused by Lys-190-Ser mutation can be complemented by *trans* supply of NS2A or NS2A_α and compensated by revertants restoring the basic nature at the NS2A_α cleavage site or by second-site changes, Asp-343 (Glu-338 in DENV) to a short, uncharged side chain in

the helicase domain of NS3. Furthermore, NS2A was found to be localized within vesicle packets (VPs), the presumed sites of RNA replication by cryoimmunogold staining (Mackenzie *et al.*, 1998). *In vitro* studies with a glutathione-S-transferase–KUN NS2A fusion protein showed that NS2A can bind to NS3, NS5, and the KUN 3'NCR (Mackenzie *et al.*, 1998). Thus, this protein is likely to be involved in coordinating the shift between RNA packaging and RNA replication, both these processes having been shown to be linked (Khromykh *et al.*, 2001b).

1.2.4.3 NS2B

NS2B (14 kDa) is a small membrane-associated protein containing a conserved central hydrophilic region flanked by two hydrophobic domains which mediates membrane association. It forms a stable complex with the NS3 protease domain and acts as a cofactor necessary for activation of NS3 protease activity (Chambers *et al.*, 1991; Chambers *et al.*, 1993; Falgout *et al.*, 1991; Yusof *et al.*, 2000). The central hydrophilic domain (~ 40 residues) within NS2B is essential for protease activation (Falgout *et al.*, 1993; Leung *et al.*, 2001). Structures of DENV2 NS2B₄₇/NS3 protease domain (NS3pro) apo enzyme and WNV NS2B₄₇/NS3pro in complex with a protease inhibitor have been determined recently, both of which contain the 47 amino acids of the central hydrophilic region of NS2B. The structural details and functional insights of NS2B/NS3 protein complex will be discussed in the next section.

1.2.4.4 NS3

NS3 (68 kDa) is a large multifunctional protein endowed with protease, helicase, NTPase, as well as 5'-terminal RNA triphosphatase activities and plays important roles in viral polyprotein processing, genome replication and virus particle assembly (Lindenbach *et al.*, 2007). Apart from its enzymatic activities, NS3 is also involved in virus assembly and infectious viral particle production (Kummerer and Rice, 2002; Liu *et al.*, 2002; Patkar and Kuhn, 2008).

1.2.4.4.1 NS3 protease domain (NS3pro)

The N terminal (aa 1-180) of NS3 encodes a trypsin-like serine protease with NS2B acting as a cofactor (Chambers *et al.*, 1991; Falgout *et al.*, 1991; Wengler *et al.*, 1991; Yusof *et al.*, 2000). The conserved serine protease catalytic triad residues are His-51, Asp-75, and Ser-135 in DENV2 (Valle and Falgout, 1998). NS2B/NS3pro complex is pivotal in viral polyprotein processing, responsible for all cleavage between or within the viral proteins at the cytoplasmic side of ER membrane, including NS2A/NS2B, NS2B/NS3, NS3/NS4A, NS4A/NS4B, NS4B/NS5 junctions, and within C, NS2A, NS4A and NS3 itself (**Figure 1. 5**) (Lindenbach *et al.*, 2007).

Except for the NS2B/NS3 junction, which contains a glutamine residue at the P2 position, the NS2B/NS3pro complex preferentially cleaves a pair of dibasic amino acids (Arg or Lys) at the P1 and P2 positions, followed by a short side chain amino acid (Gly, Ala, or Ser) at the P1' site (Li *et al.*, 2005). This is likely to be true for NS3 proteases from all four serotypes DENV1-4 (65-70% amino acid sequence identity). They share very similar, if not identical, peptide substrate structure activity relationships and it is thus possible to develop a single inhibitory agent targeting all four dengue NS3 proteases (**Figure 1. 9**) (Li *et al.*, 2005).

Figure 1. 9 Comparison of subsite specificities of NS3 proteases with their natural cleavage sites in viral polypeptide. (A) Summary of the P4-P4' substrate specificities of the NS3 proteases. Single letter amino acid abbreviations are used above and below the schematic. (B) The P6-P6' boundary cleavage sites between

NS2A and NS2B (2A/2B), NS2B and NS3 (2B/3), NS3 and NS4A (3/4A), and NS4B and NS5 (4B/5) are listed. DENV1, Singapore strain; DENV2, New Guinea C strain; DENV3, H87 strain; D, Dominica strain. The P1 and P1' in each site is separated by a space to indicate the scissile bond. Adapted from Li *et al.*, 2005.

The crystal structure of DENV2 NS2B₄₇NS3pro reveals a chymotrypsin-like fold with two β -barrels, each formed by six β -strands, with NS2B contributing an additional β -strand (residues 51-57 of NS2B) to the N-terminal β -barrel (**Figure 1. 10**) (Erbel *et al.*, 2006). The conserved catalytic triad (His-57, Asp-75 and Ser-135) is located at the cleft formed by the two β -barrels. The same topological organization applies to WNV NS2B/NS3pro. In both cases, 47 residues from the central hydrophilic region of NS2B (NS2B₄₇) were fused to NS3pro via a flexible Gly₄SerGly₄ peptide linker to generate single chain active NS2B/NS3 protease (Erbel *et al.*, 2006). The WNV NS2B₄₇NS3pro structure was determined in complex with the Bzl-Nle-Lys-Arg-Arg-H protease inhibitor, C-terminal part of NS2B encircled the C-terminal β -barrel of NS3pro by making contact with β -strands B2a, B2b, and a loop between E1b and F1. Notably, residues Gly-83 to Gln-86 of NS2B are directly involved in the formation of S2 and S3 pockets for inhibitor (substrate) binding (**Figure 1. 10**). A similar construct containing only the N terminal half of NS2B cofactor (residues 49-66) yielded soluble but catalytically inactive protein, which could not bind substrate-based inhibitor as shown by NMR spectroscopy (Erbel *et al.*, 2006). Thus, it is clear that NS2B not only stabilizes the NS3 protease by providing additional β -strands, it also completes the substrate-binding site with its C-terminal region. This differs substantially from HCV NS4A/NS3pro, in which NS4A activates NS3pro by providing a more rigid and precise framework for residues that form the substrate binding channel (Kim *et al.*, 1996; Love *et al.*, 1996).

Figure 1. 10 Structures of DENV2 NS2B/NS3pro and WNV NS2B/NS3pro with an inhibitor. NS2B and NS3pro are yellow and grey respectively. Secondary structural elements are numbered. Catalytic triad is depicted as grey sticks. Inhibitor (Bzl-Nle-Lys-Arg-Arg-H) is depicted as orange sticks. Catalytic triad is depicted as grey sticks. Adapted from Erbel *et al.*, 2006.

1.2.4.4.2 NS3 helicase domain (NS3hel)

The C-terminal domain (residues 170-618) of the flavivirus NS3 is a DEAH RNA helicase belonging to SF2, according to the classification of helicases based on the presence of certain conserved sequence signature motifs (Gorbalenya and Koonin, 1993; Tanner and Linder, 2001; Xu *et al.*, 2006). Besides DENV NS3, the helicase/NTPase activities and structures of NS3 protein from several members of the flavivirus, including WNV, YFV, JEV and MVEV, have been studied (Lescar *et al.*, 2008). The flavivirus NS3 helicase domain contains (1) RNA helicase activity required for unwinding duplex RNA and/or RNA secondary structures during genome replication by NS5 RdRp; (2) Nucleoside 5'-triphosphatase (NTPase) activity which is required to provide energy for the helicase unwinding activity; (3) 5'-terminal RNA triphosphatase activity (RTPase) which plays a role in the synthesis of the 5'-cap structure; (4) A newly proposed role in virus assembly which is independent of its known enzymatic functions (Benarroch *et al.*, 2004; Borowski *et al.*, 2000; Chambers *et al.*, 1993; Lescar *et al.*, 2008; Li *et al.*, 1999; Patkar and Kuhn, 2008; Wengler *et al.*, 1991). Dengue viruses and bovine viral diarrhea virus with impaired helicase activity

are not able to replicate, demonstrating the importance of NS3 in the *Flaviviridae* life cycle (Grassmann *et al.*, 1999; Matusan *et al.*, 2001).

NTPase activity involves binding of NTP, hydrolysis of NTP at its γ phosphoric anhydride bond and releasing NDP and inorganic phosphate. Being common for many helicases and motor proteins, the conserved motifs directly involved in NTP hydrolysis are motif I (as known as Walker A or P-loop, involved in binding to the nucleotide phosphates), motif II (as known as Walker B, involved in coordination with the divalent metal ion - Mg^{2+} or Mn^{2+} and activation of the attacking water), and motif VI (involved in binding the nucleotide phosphates) (Benarroch *et al.*, 2004; Enemark and Joshua-Tor, 2008; Li *et al.*, 1999; Wu *et al.*, 2005; Xu *et al.*, 2006). The NTPase activity is very basal and can be stimulated by ssRNA, like polyU, up to 100 times (Li *et al.*, 1999; Wang *et al.*, 2009; Xu *et al.*, 2006). This RNA stimulated NTPase activity is thought to be due to an allosteric effect, where conformation change of protein upon RNA binding results in a more kinetically favorable active site for NTP interaction (Preugschat *et al.*, 1996). NS5 can stimulate NTPase activity in a dose-dependent manner up to the point of 1:1 molar ratio (Yon *et al.*, 2005).

Helicase activity involves separation of duplex oligonucleotides into single strands utilizing the energy released from NTP hydrolysis. Flavivirus NS3hel has a 3' to 5' directionality and can unwind dsRNA with 3' overhang (Benarroch *et al.*, 2004; Li *et al.*, 1999; Xu *et al.*, 2005). Unlike HCV NS3, flavivirus NS3 unwinds dsDNA very inefficiently (Chernov *et al.*, 2008; Frick *et al.*, 2004; Pang *et al.*, 2002; Tai *et al.*, 1996; Wang *et al.*, 2009; Xu *et al.*, 2005). In contrast to HCV NS3 and Vaccinia virus NPH-II (another SF2 RNA helicase) (Gross and Shuman, 1996; Jankowsky *et al.*, 2000; Pang *et al.*, 2002), flavivirus NS3 helicase does not appear to be a processive RNA helicase. The unwinding ability decreased dramatically as the duplex region extends from 22 or 23 to 36 basepairs (Wang *et al.*, 2009), and ATP hydrolysis is absolutely required (Benarroch *et al.*, 2004; Li *et al.*, 1999; Wang *et al.*, 2009). On the other hand, for helicase assays using dsRNA substrate with a short duplex region (less than 20 bp), it is shown that the NTPase and helicase activities can be functionally uncoupled (Borowski *et al.*, 2002; Matusan *et al.*, 2001; Sampath *et al.*, 2006). The RNA helicase activity of flavivirus is also different from that of DEAD box RNA helicases too. The DEAD box helicases usually demonstrate the poorest RNA

unwinding activity and they need assistance of cofactors (Bleichert and Baserga, 2007; Cordin *et al.*, 2006; Jankowsky and Fairman, 2007; Pyle, 2008). Thus, it is not known for certain how the energy from ATP hydrolysis is utilized for unwinding dsRNA and there are probably differences in coupling the two enzymatic activities among RNA helicases of different groups (Pyle, 2008).

RNA triphosphatase (RTPase) activity involves hydrolysis of the γ phosphoric anhydride bond of triphosphorylated RNA which is the first of the three sequential enzymatic reactions for addition of 5' cap to RNA (Dong *et al.*, 2008; Lindenbach *et al.*, 2007). Mutagenesis studies and competition experiments suggest that RTPase and NTPase activities share a common active site in Flavivirus NS3 (Bartelma and Padmanabhan, 2002; Benarroch *et al.*, 2004; Wang *et al.*, 2009). Unsurprisingly, an exposed 5'-terminus makes the 5'-triphosphate group more accessible to the active site, and thus a better substrate for the RTPase activity (Wang *et al.*, 2009). The RTPase activity is likely to be divalent metal ion dependent (Benarroch *et al.*, 2004; Wang *et al.*, 2009), although divalent metal ion independence was also reported (Bartelma and Padmanabhan, 2002). Furthermore, recent studies show that NS5 can also stimulate RTPase activity in a dose-dependent manner up to the point of 1:1 molar ratio (Yon *et al.*, 2005).

Crystal structures were first reported for the NS3 helicase domains from DENV2 and YFV (Wu *et al.*, 2005; Xu *et al.*, 2005). Subsequently, a number of NS3 structures for other flavivirus helicases have been published (Lescar *et al.*, 2008; Mancini *et al.*, 2007; Mastrangelo *et al.*, 2006; Speroni *et al.*, 2008; Yamashita *et al.*, 2008). Each of these proteins comprises three subdomains. All known sequence motifs identified for the members of the SF2 superfamily of helicases are located within Rec-A like subdomains 1 and 2 (Lescar *et al.*, 2008). The ATPase active site is housed between these two subdomains and is structurally well conserved compared to other SF2 helicases including the one from HCV (**Figure 1. 11**) (Xu *et al.*, 2005; Yao *et al.*, 1997). There is no functionally meaningful structure of NS3 in a nucleotide complex currently available, although a structure of YFV NS3hel in a ADP complex (with very high B factor for the ligand; in the absence of the catalytic divalent metal ion) is reported (Wu *et al.*, 2005). It is generally proposed that the ATP is primarily held through its triphosphate moiety via contacts mediated through the divalent metal ion and the residues from the motif I, II and motif VI. Unlike DEAD box RNA

helicases, there is no Q motif found upstream of motif I, which is consistent with the observation of no discrimination between various nucleotides as substrate (Benarroch *et al.*, 2004; Locatelli *et al.*, 2002). Like the HCV NS3, a wide single stranded nucleic acid binding tunnel is formed between subdomain 1, 2 and 3, although the folding of subdomain 3 is most divergent between HCV and flavivirus NS3 (Xu *et al.*, 2005). The conserved motifs involved in ssRNA binding and duplex unwinding spread on top of the two RecA-like domains facing towards the nucleic acid binding tunnel (**Figure 1. 11**).

Figure 1. 11 Structure of DENV2 NS3 helicase. (A) Overall fold of DENV2 NS3 helicase (171-618). Secondary structural elements are labeled. The same color code

(and the relative orientation between the three domains) is also used in the topology diagram. The sulfate ion bound in the ATPase catalytic site is represented as sticks. (B) Folding diagram of the DENV2 NS3 helicase (171-618) using the same color code. The location of the conserved SF2 helicases motifs is indicated. The missing segment connecting strands β 2A and β 3 in domain I comprising residues 244 to 251 is drawn as a broken line. From Xu *et al.*, 2005.

1.2.4.4.3 Full length NS3 (NS3FL)

Although the protease and the helicase domain of NS3 are able to carry out their own function independent of each other, the full length NS3 exhibits superior helicase activity compared to the helicase domain alone (Xu *et al.*, 2005; Yon *et al.*, 2005). The active catalytic triad is dispensible for stimulatory effect suggesting that the overall folding of the N-terminal protease domain contributes to this enhancement (Yon *et al.*, 2005). On the other hand, the helicase domain has higher ATPase activity than the full length DENV2 NS3 (Xu *et al.*, 2005). Interestingly, the full-length NS3 enzymes also undergoes auto-cleavage in *cis* within its helicase domain (Bera *et al.*, 2007). Thus, the N-terminal protease domain of dengue virus NS3 has a strong influence on its helicase and ATPase activity. As a multi-functional enzyme, NS3 works together with NS2B for viral polyprotein processing and with NS5 as part of the viral replication complex (Erbel *et al.*, 2006; Johansson *et al.*, 2001).

The crystal structure of an engineered HCV protein containing the complete NS3 and the protease activation fragment of non-structural protein 4A (NS4A) reveals that helicase and protease domains are segregated and connected by a single strand peptide linker (PDB: 1CU1) (Yao *et al.*, 1999). The C-terminal end of NS3 from HCV is inserted into the active site of the protease domain mimicking the *cis* cleavage that occurs at the NS3-NS4A junction (Yao *et al.*, 1999). A model of DENV NS3FL was generated based on the topological arrangement of the two domains in HCV NS3FL (**Figure 1. 12**)(Xu, 2007). This model would at least require the inter-domain linker (residues 169-179 in NS3FL) to rearrange dramatically as the distance between C terminus of the protease and the N terminus of the helicase is 59 Å (Xu, 2007). Nevertheless, the model does not provide significant information concerning the

stimulatory effect on helicase activity. Thus, further structural studies of flavivirus full length NS3 are needed.

Figure 1.12 Model of dengue full length NS3. The model is generated based on superposition of structures of dengue NS3 helicase (PDB: 2BMF) and protease (PDB: 2FOM) domains to HCV NS3FL-NS4A structure (PDB: 1CU1) (Erbel *et al.*, 2006; Xu *et al.*, 2005; Yao *et al.*, 1997). DEN2 NS3 helicase domain starts from amino acid 171 to 618, with three extra amino acids (Ala-Met-Ala) introduced by NcoI at the N terminus. The protease domain starts from 18 to 167 (Erbel *et al.*, 2006). Distance between the N-terminus of helicase domain and the C-terminus of protease domain is 59 Å. Dengue NS3 helicase, protease and NS2B cofactor are in blue, purple and yellow, respectively. The catalytic triad (His-51, Asp-75, Ser-135) is depicted as cyan sticks. From Xu, 2007.

1.2.4.5 NS4A

NS4A (16 kDa) is a small hydrophobic protein and its function still remains elusive (Lindenbach *et al.*, 2007). The N-terminus NS4A is cleaved by the viral NS2B-NS3 protease in the cytoplasm. The C-terminal 23 residues signal sequence, also called the 2K fragment is removed from the N-terminus of NS4B by host signalase in the ER lumen, after viral NS2B-NS3 protease cleavage at the 4A/2K site (Lin *et al.*, 1993a; Roosendaal *et al.*, 2006). Four membrane associated sequences predict that DENV NS4A is an integral membrane protein (**Figure 1.13**) (Miller *et al.*,

2007). NS4A is able to form homo oligomers thus inducing distinct membrane alternations, associates with other NS proteins, and colocalizes with dsRNA (Lindenbach and Rice, 1999; Mackenzie *et al.*, 1998; Miller *et al.*, 2007). Localization of NS4A into VPs (the presumed sites of viral RNA replication) as well as into CMs and PCs (the place for viral polyprotein processing) further implies that NS4A plays important roles in viral RNA replication (Lin *et al.*, 1993a; Lindenbach and Rice, 1999; Mackenzie *et al.*, 1998; Miller *et al.*, 2007; Roosendaal *et al.*, 2006; Westaway *et al.*, 2003). Further studies will be required to clearly define the contributions of NS4A in inducing membrane rearrangements and viral replication.

Figure 1. 13 Model for the membrane topology of DENV NS4A. The N-terminal one-third of DV NS4A is localized in the cytoplasm; pTMS 1 (potential transmembrane segment) and pTMS 4 (2K fragment) span the membrane from the cytoplasmic to the luminal site, whereas pTMS 3 seems to span the lipid bilayer from the luminal to the cytoplasmic site. Amino acids around positions 76–99 (pTMS 2) most probably do not span the membrane but are closely associated with the luminal side of the lipid bilayer. From Miller *et al.*, 2007.

1.2.4.6 NS4B

NS4B (27 kDa) is the largest hydrophobic NS protein of flavivirus and its function is not clearly defined (Lindenbach *et al.*, 2007). The N-terminus of NS4B is processed from 2K fragment of NS4A by host signalase in ER and the NS4B-NS5 junction is cleaved by the viral NS2B-NS3 protease in the cytoplasm (Clum *et al.*, 1997; Falgout *et al.*, 1991; Roosendaal *et al.*, 2006). NS4B is an integral transmembrane protein independent of the 2K signaling peptide from NS4A. The membrane topology of NS4B was determined (**Figure 1. 14**) (Miller *et al.*, 2006).

NS4B colocalizes with NS3 and dsRNA in cytoplasmic foci originating from ER, arguing that NS4B is part of the membrane-bound viral replication complex (Miller *et al.*, 2006). The direct interaction of NS4B with NS3 helicase domain was identified by using yeast two-hybrid, pull-down and immunoprecipitation assays (Umareddy *et al.*, 2006). NS4B is also able to dissociate NS3 from ssRNA and consequently enhance the helicase activity of NS3 in an in vitro unwinding assay (Umareddy *et al.*, 2006). These findings suggest NS4B modulates viral replication via its interaction with NS3 and dsRNA.

Figure 1. 14 Model for the membrane topology of DENV 2K-NS4B. 2K serves as a signal sequence for the translocation of the N terminus of NS4B into the lumen of the ER. The predicted TMDs 1 and 2 do not span the membrane but rather reside in the ER lumen. TMDs 3, 4, and 5 span the membrane and serve as internal signal sequences for membrane association. Glycosylation may occur at an internal N-glycosylation site in the predicted TMD 2. From Miller *et al.*, 2006.

1.2.4.7 NS5

NS5 (103kDa) is the largest multifunctional protein encoded by the viral genome (Lindenbach *et al.*, 2007). The N-terminal domain of NS5 is an S-adenosylmethionine dependent methyltransferase (MTase) involved in 5' capping of the viral genome (Dong *et al.*, 2008; Koonin and Dolja, 1993). The C-terminal domain possesses RNA dependent RNA polymerase (RdRp) activity, essential for

viral replication (Rawlinson *et al.*, 2006; Tan *et al.*, 1996). Residues 320 to 405 contain two nuclear localization signals (β 1 NLS and α/β NLS), which might play important role for transportation of NS5 into nucleus and direct the interaction between NS5 and NS3 (**Figure 1. 15**) (Brooks *et al.*, 2002; Forwood *et al.*, 1999; Johansson *et al.*, 2001).

Figure 1. 15 Schematic representation of the distribution of the functional domains in the DENV NS5 protein. The FL NS5 protein has three major functional domains. The N-terminal MTase spans residues 1 to 296. The NLS has been divided into α/β -NLS or β -NLS. The β -NLS region is thought to interact with the NS3 viral helicase. The six conserved amino acid sequence motifs within the RdRp domain (residues 273-900) are colored respectively and P denotes the priming loop site. From Yap *et al.*, 2007.

1.2.4.7.1 S-adenosylmethionine dependent methyltransferase (MTase)

MTase is essential for viral replication and is responsible for the sequential methylation on N-7 of Guanosine-0 base and then on 2'-O of the ribose of Adenosine-1 of the viral RNA cap (type I cap structure: m7G⁰pppA¹m-RNA) (Dong *et al.*, 2008; Wengler, 1981). The two steps of methylation require specific RNA elements at the 5'UTR region of the viral genome, while non-viral originated capped RNA and dinucleotide cap analogues are not active substrates (Dong *et al.*, 2007; Ray *et al.*, 2006).

Several crystal structures of flavivirus NS5 MTase domain have been determined (Assenberg *et al.*, 2007; Egloff *et al.*, 2002; Mastrangelo *et al.*, 2007a; Zhou *et al.*, 2007). They adopt a similar globular fold with three subdomains (**Figure 1. 16**). The core subdomain 2 folds similarly to the catalytic domain of other AdoMet-dependent MTases and is flanked by two extensions (Egloff *et al.*, 2002). Within MTase, these conserved active sites, namely, (1) SAM (S-adenosyl-l-methionine, methyl donor) / SAH (S-adenosyl-l-homocysteine, by-product of methyl transfer reaction) binding site, (2) GTP binding site, and (3) putative RNA binding site, have

been mapped by both crystallography and structure based mutagenesis studies (**Figure 1. 16**) (Dong *et al.*, 2008). A substrate repositioning model for flavivirus RNA cap methylation is proposed, in which translocation of the RNA is required in order to complete the two methylation reactions. However, whether the sequential methylation process is carried out by a single or two MTase molecules is still unknown (Dong *et al.*, 2008).

In addition, one missing step preceding the MTase reaction is RNA guanylyl transfer reaction ($\text{pppG} + \text{ppA-RNA} \rightarrow \text{GpppA-RNA}$). Based on structural and biochemical studies on MTase substrate interaction, Egloff *et al.* hypothesize that MTase also possesses guanylyltransferase (GTase) activity (Egloff *et al.*, 2007). Experiments are required to demonstrate such an enzymatic activity.

Structures of the MTase in complex with viral RNA are awaited to help verify the model and to answer these questions.

Figure 1. 16 Structure of the flavivirus MTase. (A) Ribbon representation of the crystal structure of the WNV MTase. Three domains of the MTase structure are: N-terminal domain, red; MTase core, green; and C-terminal domain, cyan. The bound SAH (S-adenosyl-l-homocysteine, by-product of methyl transfer reaction) molecule is shown in a ball-and-stick representation, with atom colors as follow: carbon, yellow; oxygen, red; nitrogen, blue; and sulfur, green. (B) Three flavivirus-conserved binding sites for SAH/SAM, GTP, and RNA. The binding sites are shown on the surface of WNV MTase, which in this panel is depicted as an electrostatic potential map: positively charged residues in blue, negatively charged in red, and neutral in white. The SAH/SAM-binding pocket is identified based on the co-crystal structure of the WNV MTase with SAH. The GTP-binding site of the WNV MTase was modeled

through structural alignment with the DENV2 GTP–SAH–MTase tertiary complex. The RNA binding site is indicated by the positively charged residues. Adapted from Dong *et al.*, 2008.

1.2.4.7.2 RNA dependent RNA polymerase (RdRp)

The C-terminal region of NS5 possesses RNA dependent RNA polymerase (RdRp) activity. Six conserved motifs A to F are characteristics of polymerases from positive-strand RNA viruses (**Figure 1. 15**) (Koonin, 1991; Rice *et al.*, 1985). Together with other viral proteins and poorly identified host cell proteins, the NS5 polymerase domain is responsible for the initial synthesis of (-) anti genome RNA forming a transient double stranded RNA replicative intermediate (RI). Using the newly synthesized (-) anti genome RNA as template, the RdRp synthesizes additional (+) genomic RNA (Kapoor *et al.*, 1995; Nomaguchi *et al.*, 2003; You *et al.*, 2001).

The crystal structure of DENV3 RdRp domain (residues 273-900) was determined recently (Yap *et al.*, 2007) (**Figure 1. 17 A/B**). The structure assumes a canonical right hand conformation consisting of fingers, palm and thumb, characteristic of known polymerase structures; the NLS region is distributed between the fingers and thumb subdomains, important in stabilizing the structure and regulating the activity of the RdRp (Kohlstaedt *et al.*, 1992; Yap *et al.*, 2007). The DENV RdRp catalytic domain can be superimposed to the WNV polymerase with a r.m.s. deviation of 1.23 Å for 313 equivalent Cα and there is little conformational change occurs within the individual subdomains (Malet *et al.*, 2007). Interestingly, the DENV 3 RdRp assumes a more “open” conformation compared to the conformation observed for WNV RdRp, with the fingers subdomain rotating away from the thumb by an angle of approximately 8° (Yap *et al.*, 2007).

The N terminal fingers subdomain and especially the “fingertips region” appear quite mobile compared to the other palm and thumb subdomains. The concave surface formed by solvent-exposed residues from the base of the fingers domain was suggested to be involved in accommodating the N-terminal MTase domain. The palm domain appears the most structurally conserved amongst known polymerases, reflecting the preservation of architecture of the catalytic site during evolution. Four out of six conserved sequence motifs responsible for NTP binding and catalysis are

found in the palm domain, including the motif C (GDD catalytic active site comprising Asp-663 and Asp-664). The thumb domain at the C terminal end of the RdRp of DENV is the structurally most variable amongst known polymerase structures. Motif E forms an anti-parallel β sheet wedged between the palm domain and several α helices of the thumb domain. A second loop spanning amino-acids 782-809 forms the priming loop that partially occludes the active site, participating in rNTP binding and *de novo* initiation. Two zinc binding pockets are found in the thumb and fingers subdomains respectively, which are proposed to contribute to structural stability and conformational switching of the protein (Yap *et al.*, 2007).

The relative spatial organization of the MTase and RdRp domains in *Flavivirus* NS5 is unknown. However, a model of full-length NS5 was proposed based on reverse genetic experiments on DENV2 NS5 and structural information on DENV MTase (PDB code: 1L9K) and WNV NS5 RdRp (PDB code: 2HFZ) (Malet *et al.*, 2007). In the model, the RNA-substrate binding groove of the MTase domain is positioned near the RdRp dsRNA exit tunnel, structurally and mechanistically consistent with RNA cap methylation of the nascent genome occurring after the (+) RNA product leaves the RdRp domain (**Figure 1. 17 C**) (Malet *et al.*, 2007).

Figure 1. 17 Structure representation of DENV 3 RdRp. (A) Ribbon representation of the overall structure of DENV 3 RdRp (residues 273-900). The various subdomains are as follows: Fingers (blue), NLS (yellow), Palm (green), and Thumb (red). Secondary structure elements as defined in Fig 3 are labeled. The N- and C-termini are displayed and the residues at the boundaries of the missing regions in the model are numbered. (B) The six conserved motifs present in various RdRps structures are mapped onto a ribbon representation of the DENV 3 RdRp catalytic domain and as follows: motif A (dark green), motif B (green), motif C (red), motif D (dark blue), motif E (magenta), motif F (light blue) and priming loop (pink). Asp-663 and Asp-664 from the motif C of the catalytic site are shown in sticks representation and labeled. The divalent metal ions Zn^{2+} and Mg^{2+} are represented by spheres in gold and purple respectively. Adapted from (Yap *et al.*, 2007). (C) Flavivirus full length NS5 model. Top view of the WNV RdRp (pink) and the MTase (cyan) domains after docking. Interaction interfaces are highlighted and the key subdomains and motifs are labeled. Adapted from Malte *et al.*, 2007.

1.3 RNA helicases: a structural view

RNAs are of the most structurally and functionally diverse biomolecules. They are well known as the messenger of genetic information, i.e., mRNA of prokaryotic and eukaryotic organisms and virus genomic RNA. RNA is also an important

structural component of macromolecular complexes. Examples of ribonucleoprotein complexes (RNPs) include ribosome and mRNA spliceosome, where RNA is the catalyst. Because of the great diversity and flexibility of RNA molecules, the hypothesis of a RNA world proposes that life based on RNA predated the current life based on DNA and protein (Ganem, 1987; Meyers *et al.*, 2008; North, 1987). To achieve the various functions, RNA must correctly fold into the proper tertiary structure and associate with the correct set of proteins.

“Helper” proteins such as RNA helicases that can unwind dsRNA or RNA/DNA hybrid, dissociate proteins from RNA molecules, resolve RNA secondary structures and assist in proper tertiary structure formation similar to protein chaperones during protein folding. Hence, initially seen as proteins unwinding RNA duplexes by utilizing the energy of NTP hydrolysis, RNA helicases are also known as RNA chaperones or RNPases. RNA helicases are involved in a variety of biological processes including transcription, ribosome biogenesis, pre-mRNA editing and splicing, RNA export to the cytoplasm, translation initiation and termination, RNA degradation, organelle gene expression, and virus propagation (Rajkowitsch *et al.*, 2007; Tanner and Linder, 2001). Being indispensable in almost all aspects of RNA metabolism, RNA helicases have been subjected to intense genetic, biochemical, structural, biophysical, and bioinformatic studies over the last decades (Bleichert and Baserga, 2007; Jankowsky and Fairman, 2007; Pyle, 2008; Rajkowitsch *et al.*, 2007).

In this section, recent advances in structural biology of RNA helicases (**Figure 1. 18**) are discussed in order to gain molecular insights into the mechanisms of RNA remodeling (RNA binding, translocating along ssRNA, and unwinding dsRNA, etc.) and NTP binding/hydrolysis, the coupling between the two activities and the regulation by the related cofactors.

1.3.1 Classification of RNA helicases

Gorbalenya and Koonin grouped helicases into five different superfamilies, designated SF1–5, based on conserved sequence motifs. SF1 and SF2 are the two largest and most closely related families of helicases. They contain a large number of DNA and RNA helicases from archaea, eubacteria, eukaryotes and viruses (Caruthers and McKay, 2002; Gorbalenya and Koonin, 1993). Sequence analysis revealed that

members of SF1 and SF2 share at least seven conserved motifs (I, Ia, II-VI) (Hall and Matson, 1999). A new motif, Q motif, which is upstream of motif I, was discovered in the DEAD-box family of helicases (Tanner *et al.*, 2003). The conserved motifs are crucial for various helicase functionalities, i.e., ATP binding and hydrolysis (motif I, II, and VI), RNA binding (motif Ia, IV, and V), and the coupling between ATP hydrolysis and helicase activity (motif III). SF3, 4 and 5, sharing less conserved motifs and structural similarities in between and with SF1 and SF2, are usually hexameric helicases, some of which are more similar to the AAA/AAA+ protein families (Caruthers and McKay, 2002; Singleton *et al.*, 2007).

There are other means of grouping helicases, according to their functional, structural and mechanical properties e.g., substrate preference (DNA vs RNA), processivity (processive vs nonprocessive), directionality (3' to 5' vs 5' to 3') and oligomerization (monomeric vs hexameric).

Most RNA helicases (**Figure 1. 18**) belong to the SF2 superfamily and are grouped into DEAD box, the DEAH and the DExH subfamilies, which are generally referred to as DExD/H box RNA helicases, named after motif II (Caruthers and McKay, 2002; Hall and Matson, 1999; Silverman *et al.*, 2003; Tanner and Linder, 2001). A few RNA helicases, the Upf1-like helicases, belong to SF1 (Cheng *et al.*, 2007b; Jankowsky and Fairman, 2007). Together with the DNA helicases in SF1 and SF2, they constitute the largest number of helicases that are sequentially and structurally related but functionally diversified.

Figure 1.18 Sequence and structure organization of RNA helicases.

(A) Sequence characteristics of SF2 and SF1 helicases. The dendrogram (left) indicates the phylogenetic relationship between the SF1 and SF2 helicase families. Subgroups containing RNA helicases are in bold. Helicase domains are represented as dark gray blocks, and C and N termini as light gray blocks. Insertions in the helicase domains are labeled. The length of the blocks is not to scale. Conserved sequence motifs are colored according to their biochemical function: red, ATP binding and hydrolysis; yellow, coordination between polynucleotide binding and ATPase activity; blue, nucleic acid binding. The name of the subgroup derives from the sequence of motif II, in single-letter code, although the nature of all characteristic sequence motifs in a given protein determines to which subfamily it belongs. (B) Topology of the two RecA-like helicase domains. Elements with solid outlines are present in all SF1 and SF2 structures; elements with dashed outlines are not present in all proteins. The position of the conserved sequence motifs is indicated by numbered octagons, colored as in (A). Domains 1B and 1C of the Upf1 group are inserted before motif I and in between motifs Ib and Ic. Adapted from Jankowsky and Fairman, 2007.

1.3.2 Conserved motifs

Based on the structural, biochemical and genetical data, the proposed sequential and functional features of each conserved motif is summarized in this section. In total, there are seven to ten motifs identified, among which seven motifs (I, Ia, II, III, IV, V and VI) are conserved in almost all SF1 and SF2 helicases and are better characterized (**Figure 1. 19**) (Gorbalenya and Koonin, 1993). On the other hand, motifs Ib, Ic, IVa and Q are either less conserved or absent in some helicases (Cordin *et al.*, 2006; Gorbalenya and Koonin, 1993; Jankowsky and Fairman, 2007).

Figure 1. 19 Conserved motifs in the DEAD box and related DExD/H box helicase families. Sequences of the conserved motifs from the yeast *S. cerevisiae* eIF4A (DEAD box helicase), Prp2 (DEAH), NS3 (DExH box helicase from the hepatitis C virus) and Ski2 (DExH). Symbols used are as follow: o: S,T;l: I, L, V;x: any residue;a: F, W, Y;c: D, E, H, K, R;h: A, F, G, I, L, M, P, V, W, Y;+: H, K, R;u: A, G. Adapted from Cordin *et al.*, 2006.

1.3.2.1 Motifs involved in ATP binding and hydrolysis

Motif I, also known as the Walker A motif or P-loop (phosphate binding loop), is common to all helicases and many non-helicase NTPases in general (Gorbalenya *et al.*, 1988; Walker *et al.*, 1982). In SF1 and SF2, motif I (AxTGoGKT) is structurally equivalent and its sequence is highly conserved (Cordin *et al.*, 2006). It is directly involved in NTP binding by interacting with the phosphates of the nucleotide and with the coordinated divalent metal ions (Mg^{2+} or Mn^{2+}) (Caruthers and McKay, 2002; Shi *et al.*, 2004; Xu *et al.*, 2005). The conformation of the P-loop seems to be flexible and is stabilized by its ligands. Within the exon junction complex, the P-loop of

eIF4AIII (a member of DEAD box protein) adopts a uniform conformation in order to accommodate AMPPNP-Mg²⁺ (Andersen *et al.*, 2006; Bono *et al.*, 2006). Instead of being flexible in the apo enzyme, the side chain of Lys-88 (Motif I) forms stable H bonds with both β and γ phosphates (Andersen *et al.*, 2006; Bono *et al.*, 2006). The P-loop of UAP56 (another member of DEAD box helicase) undergoes an opening transition upon ADP-Mg²⁺ binding (PDB codes: 1XTI for apo enzyme; 1XYJ for ADP-Mg²⁺ complex) (Shi *et al.*, 2004). In the absence of ADP-Mg²⁺, Lys-95 (motif I) contacts Asp-196 (motif II). This interaction is disrupted and protein-ligand interactions are formed (Mg²⁺-Asp-196 and β -phosphate-Lys-95) upon the binding of ADP-Mg²⁺ (Shi *et al.*, 2004). The available crystal structures of the HCV NS3 helicase (DECH helicase) suggests that P-loop adopts a more rigid conformation with no obvious change in the presence (PDB codes: 1AIV, 1CUI, and 2F55) or absence of SO₄ (PDB codes: 1HEI and 2OHM) (Cho *et al.*, 1998; Kim *et al.*, 1998; Mackintosh *et al.*, 2006; Yao *et al.*, 1997; Yao *et al.*, 1999). There is no direct interaction between Lys-209 (motif I) and Asp-289 (motif II) in any of the structures. ssDNA binding does not affect the P-loop conformation either (Kim *et al.*, 1998; Mackintosh *et al.*, 2006). The P-loop of Flavivirus NS3 helicases (DEAH helicase) (PDB codes: 1YKS, 1YMF, 2BMF, 2QEQ, 2V6I, 2V8O, and 2Z83) behaves similarly to that of the HCV NS3 helicase (Mancini *et al.*, 2007; Mastrangelo *et al.*, 2007c; Speroni *et al.*, 2008; Wu *et al.*, 2005; Xu *et al.*, 2005; Yamashita *et al.*, 2008). Interestingly, the presence of SO₄-Mn²⁺ in DENV2 NS3 helicase seems to stabilize the conformation of the P-loop by lowering the average B factor over 30 Å² (Xu, 2007).

Motif II, also known as the Walker B motif, is necessary for the ATPase activity (Walker *et al.*, 1982). Along with Walker A motif, Walker B is present in many families of NTPases. The consensus sequence of motif II from SF2 is **DE**xD/H, according to which the related helicases are grouped and named (Jankowsky and Fairman, 2007; Pyle, 2008; Silverman *et al.*, 2003; Tanner and Linder, 2001). It is by now clear that motif II coordinates a divalent metal ion (Mg²⁺ or Mn²⁺) in various DExD/H helicases (Enemark and Joshua-Tor, 2008; Pyle, 2008). The two conjugated acidic residues hold the ATP through the coordinated metal ion, while the glutamate residue is thought to be responsible for the hydrolysis of the β/γ phosphoanhydride bond of a bound ATP (Cordin *et al.*, 2006; Enemark and Joshua-Tor, 2008). The role

of the last residue of motif II (aspartate or histidine) is not clear, although mutations of which affect both ATPase and helicase activities of HCV NS3 helicase (Tai *et al.*, 2001). In DEAD box proteins, the last aspartate is H-bonding with the serine and threonine of motif III (Andersen *et al.*, 2006; Bono *et al.*, 2006; Cheng *et al.*, 2005; Sengoku *et al.*, 2006; Shi *et al.*, 2004). It also comes to close proximity to the histidine from motif VI when the two Rec-A like domains close upon substrate binding (Andersen *et al.*, 2006; Bono *et al.*, 2006; Sengoku *et al.*, 2006). The histidine residue of motif II in DExH helicases is close to motif III threonines as well as to motif VI glutamine (Cho *et al.*, 1998; Kim *et al.*, 1998; Mancini *et al.*, 2007; Mastrangelo *et al.*, 2007c; Speroni *et al.*, 2008; Wu *et al.*, 2005; Xu *et al.*, 2005; Yamashita *et al.*, 2008; Yao *et al.*, 1997; Yao *et al.*, 1999), except for one case of HCV NS3 helicase in which the two Rec-A like domains are in open conformation (Yao *et al.*, 1997).

Motif VI is located at the interface between domains 1 and 2 and is important for the ATPase activity (Cordin *et al.*, 2006; Enemark and Joshua-Tor, 2008; Frick, 2007). The consensus sequence is HRIGRTGRR for DEAD box helicases and QRxGRRxGRR for DEAH box helicases (Tanner and Linder, 2001). The two arginine residues highlighted above are known as “arginine fingers” which are essential for ATP binding and hydrolysis, and likely to stabilize the highly negatively charged β/γ phosphates in the course of ATP hydrolysis. Mutations on the arginine residues destroy the ATPase activity and thus the helicase activity of eIF4A (Pause *et al.*, 1993; Rogers *et al.*, 2002), HCV NS3 helicase (Kim *et al.*, 1997; Kwong *et al.*, 2000), and DENV2 NS3 helicase (Sampath *et al.*, 2006). It is noticed that the DEAD box helicases ($D^{\text{motif II}} \cdots H^{\text{motif VI}}$) and DExH box helicases ($H^{\text{motif II}} \cdots Q^{\text{motif VI}}$) have different consensus sequences in their respective motifs II and VI (Tanner and Linder, 2001). This co-variation between the motifs suggests a potential communication between the two RecA-like domains (Cordin *et al.*, 2006; Tanner and Linder, 2001).

Q motif, functioning as a molecular switch to regulate ATP binding and hydrolysis, was recently discovered and characterized for the DEAD box RNA helicases (Cordin *et al.*, 2004; Tanner *et al.*, 2003). The consensus sequence for Q motif was defined as GaccPohlQ, where a is an aromatic residue, c is a charged group, o is an alcohol, h is a hydrophobic group and l is an aliphatic residue. An additional, highly conserved aromatic residue located 17 residues (forming a conserved helix)

upstream of the Q motif was found to be associated with the motif (Tanner *et al.*, 2003). The Q motif like a cap structure on the RecA-like domain I of all DEAD box helicase structures was found to establish direct interactions with the bound nucleotide and to be necessary for specific binding of adenine. The conserved glutamine forms hydrogen bonds with the N6 and N7 positions of the adenine base and the conserved aromatic group of the Q motif stacks on the base (Andersen *et al.*, 2006; Benz *et al.*, 1999; Bono *et al.*, 2006; Sengoku *et al.*, 2006; Shi *et al.*, 2004). It was demonstrated that the Q motif is also necessary for the efficient binding of ssRNA and thus for the helicase activity of the DEAD box helicases (Cordin *et al.*, 2004). Such a Q motif is absent in the DE-x/A-H RNA helicases like Yeast Prp2, HCV NS3 and other flavivirus NS3 helicases. This is consistent with the nonspecific NTPase activity of Prp2 from Yeast and NS3 helicases from HCV and DEN (Benarroch *et al.*, 2004; Kim *et al.*, 1992; Wardell *et al.*, 1999). However, the situation for DNA helicases from SF1 and SF2 is ambiguous. SF1 DNA helicases, like PcrA (Velankar *et al.*, 1999) and UvrD (Lee and Yang, 2006) contain the conserved glutamine residue upstream of motif I which forms hydrogen bonds with the N6 and N7 positions of the adenine base; however, the aromatic residue (Tyr-286 of PcrA and Tyr-283 of UvrD) that stacks on the base is located downstream of motif I and is within the DNA-helicase specific motif IV (Lee and Yang, 2006; Velankar *et al.*, 1999). On the other hand, UvrB (SF2 DNA helicase) shows equivalent hydrogen bonding between the adenine of a bound ATP and a glutamine and it shows base stacking between the adenine and an aromatic group, both of which are upstream of motif I (Eryilmaz *et al.*, 2006; Theis *et al.*, 1999).

1.3.2.2 Motifs involved in RNA binding/remodeling

The conservation of the motifs involved in RNA binding is generally lower in comparison with that of the motifs directly involved in ATP binding and hydrolysis. This may be explained by the differences of their functionalities. ATP binding and hydrolysis is a relatively simpler task, requiring a confined chemical space and allowing a small degree of freedom. Therefore a more stringent selection of proper building elements (amino acids) and correct scaffold (binding pocket) are undertaken by the variety of NTPases (Gorbalenya and Koonin, 1993; Story and Steitz, 1992; Walker *et al.*, 1982). On the other hand, RNA binding/remodelling is a more dynamic

process involving extensive spatiotemporal rearrangements/coordination of the directly involved motifs (Enemark and Joshua-Tor, 2008; Pyle, 2008). The fact that RNA helicases are involved in almost all aspects of RNA metabolism suggests a large variety of potential modes of helicase-RNA interactions (Bleichert and Baserga, 2007; Jankowsky and Fairman, 2007; Rajkowitsch *et al.*, 2007). Meanwhile, since most of the known interactions between RNA and the protein are through the RNA backbone, common features on general chemical and structural properties are shared by the RNA binding motifs of the various RNA helicase (Caruthers and McKay, 2002; Pyle, 2008). Due to the limited number of crystal structures of RNA-helicase complexes and the relatively low extent of conservation, the motifs involved in RNA remodeling are less defined than the motifs involved in ATP binding and hydrolysis (Gorbalenya and Koonin, 1993).

Motif Ia, locating on RecA-like domain 1, are among the first seven motifs identified (Gorbalenya and Koonin, 1993). For DEAD box helicases, the consensus sequence for motif Ia is **PTRELA** (Cordin *et al.*, 2006). The structures of eIF4AIII and VASA in complex with ssRNA show that the proline and arginine residues interacts with the ribose 2-OH and phosphate from the ssRNA backbone respectively (Andersen *et al.*, 2006; Bono *et al.*, 2006; Sengoku *et al.*, 2006). The equivalent consensus sequence is **PVSAA** for HCV NS3 (Kim *et al.*, 1998) and **FTNKAA** for SF1 DNA helicases (Rep, PcrA, and UvrD) (Korolev *et al.*, 1997; Lee and Yang, 2006; Velankar *et al.*, 1999), where the interactions between the helicases and the bound ssDNA are only through the backbone phosphates. This backbone recognition feature is also commonly observed for the other nucleic acid binding motifs. The bold letters highlighted in the consensus sequences are the amino acids in direct contact with backbone phosphates, whereas the bond/underlined letters are the amino acids showing RNA specific interaction (through ribose 2-OH).

Motif Ib similar to but less conserved than motif Ia, is also located on RecA-like domain 1. It was initially referred to as the TxGx motif (Caruthers and McKay, 2002). The consensus sequence for motif Ib is **TPGR** for DEAD box helicases (Cordin *et al.*, 2006). Again, RNA specific interaction (through ribose 2-OH) was observed in the structures of eIF4AIII and VASA in complex with ssRNA (Andersen *et al.*, 2006; Bono *et al.*, 2006; Sengoku *et al.*, 2006). The amino acid sequence of motif Ib is **TYGK** for HCV NS3 (Kim *et al.*, 1998) and **TFH** for SF1 DNA helicases

(Rep, PcrA, and UvrD) (Korolev *et al.*, 1997; Lee and Yang, 2006; Velankar *et al.*, 1999).

Motif Ic (GG) is commonly present in between motifs Ia and Ib in many DEAD box helicases (Cordin *et al.*, 2006). It forms a loop structure linking $\beta 4$ and $\alpha 6$ of RecA-like domain 1 that interacts with the RNA backbone phosphate (Andersen *et al.*, 2006; Bono *et al.*, 2006; Cordin *et al.*, 2006; Sengoku *et al.*, 2006). TG is the structurally equivalent sequence in HCV NS3 and archaeal Hel308 (SF2 DNA helicase) (Buttner *et al.*, 2007; Kim *et al.*, 1998). There is no equivalent structural element reported from the available SF1 helicase structures. Instead, Lee and Yang identified another two ssDNA binding motifs from domain 1B. Domain 1B is inserted in between motif Ib and motif II from domain 1A (equivalent to RecA like domain 1 in SF2 helicases) in UvrD and other SF1 helicases as well (Lee and Yang, 2006).

Motif IV structurally mapped on RecA-like domain 2, shows poor sequence consensus among DExD/H RNA helicases (Cordin *et al.*, 2006; Tanner and Linder, 2001). In DEAD box helicases, the consensus sequence was reported as (L/V) (I/V) Fxx (T/S) x. Motif IV from HCV NS3 and archaeal Hel308 was observed with similar pattern for nucleic acid interaction (Buttner *et al.*, 2007; Kim *et al.*, 1998). It is possible that the residues within the range of motif IV can come in contact with the nucleic acid backbone in a manner that may enable corporative translocation or an unwinding process. Based on sequence and structure comparisons, motif IVa from SF1 helicases is equivalent to motif IV from DExD/H RNA helicases (Korolev *et al.*, 1998; Tanner and Linder, 2001). Motif IV from known SF1 helicase is actually involved in nucleotide binding, functionally similar to the Q motif from DEAD box helicases (Cheng *et al.*, 2007b; Korolev *et al.*, 1997; Lee and Yang, 2006; Velankar *et al.*, 1999).

Motif IVa also named as (GxxxQxxR) is a newly discovered motif based on the complex structures of the DEAD box proteins and ssRNA (Andersen *et al.*, 2006; Bono *et al.*, 2006; Sengoku *et al.*, 2006). It extends from a loop structure to the N terminal portion of $\alpha 13$, facing towards the ssRNA backbone and close to the interface between the two RecA-like domains in VASA (Sengoku *et al.*, 2006). This motif was not previously reported in any helicases (Cordin *et al.*, 2006; Tanner and Linder, 2001). Interestingly, a similar module structurally equivalent to this motif is

also present in HCV NS3 and archaeal Hel308 (Buttner *et al.*, 2007; Kim *et al.*, 1998). Similar to motif Ic, there is no equivalent structural element reported for the available SF1 helicases. Instead, another two ssDNA binding motifs from the inserted domain 2B between motif IVa and motif V was identified to be in contact with the dsDNA in UvrD and probably in other SF1 helicases as well (Lee and Yang, 2006).

1.3.2.3 Motifs linking ATPase and helicase activities

Motif III was proposed to participate in coordinating the ATPase and helicase activities (Caruthers and McKay, 2002; Cordin *et al.*, 2006; Hall and Matson, 1999; Tanner and Linder, 2001). The consensus sequence is (S/T)AT for DExD/H RNA helicases. In DEAD box helicases, motif III appears to be part of domain 1. In DExH helicases like HCV NS3 and other flavivirus NS3 proteins, it is at the hinge region between the two RecA-like domains. In both cases, motif III is next to the ATP binding pocket at the interface between the two RecA-like domains, but apparently there is no direct interaction with the γ phosphates of ATP (Kim *et al.*, 1998; Sengoku *et al.*, 2006; Wu *et al.*, 2005; Xu *et al.*, 2005). Mutations in motif III of eIF4A and in related DEAH box proteins cause dramatic loss of helicase activity but minor effects on ATP binding, ATP hydrolysis and RNA binding (Pause and Sonenberg, 1992; Rocak *et al.*, 2005; Schwer and Meszaros, 2000). Mutations in motif III of HCV NS3 severely impaired both ATPase and helicase activities (Tai *et al.*, 2001). Motif III in the SF1 helicase is composed of a different set of consensus sequence, sharing little apparent sequence similarity (Caruthers and McKay, 2002). It is in contact with both the γ phosphate nucleotide and nucleic acid (Cheng *et al.*, 2007b; Korolev *et al.*, 1997; Lee and Yang, 2006; Velankar *et al.*, 1999). Although motif III is quite different in SF1 and SF2, it is likely that motif III exhibits a similar function in helicases from both superfamilies.

Motif V similar to motif IV shows poor sequence consensus among DExD/H RNA helicases (Cordin *et al.*, 2006; Tanner and Linder, 2001). Because of the poor conservation, a longer than usual consensus sequence is defined (TDVxxRGID for DEAD box helicases) (Cordin *et al.*, 2006; Tanner and Linder, 2001). It is part of a loop at the interface between domains 1 and 2 that points towards the RNA-binding region. The most conserved threonine residue is directly H-bonded to the nucleic acid

backbone phosphate. This interaction is also observed in complex structures of nucleic acid with DExD/H helicase as well as with SF1 DNA helicases (Andersen *et al.*, 2006; Bono *et al.*, 2006; Buttner *et al.*, 2007; Kim *et al.*, 1998; Korolev *et al.*, 1997; Lee and Yang, 2006; Sengoku *et al.*, 2006; Velankar *et al.*, 1999). The C terminal region of motif V extended to the ATPase active site. There are interactions between motif V from eIF4AIII/VASA and the phosphate moiety of AMPPNP (Andersen *et al.*, 2006; Bono *et al.*, 2006; Sengoku *et al.*, 2006). However, there is no further study suggesting how the RNA binding is coupled or regulated by nucleotide binding.

1.3.3 Structure and mechanism of the RNA helicases

1.3.3.1 Overall features

All the structures of the SF1 and SF2 helicases present the same core of overall protein folding with two RecA-like domains (**Figure 1. 20**) (Caruthers and McKay, 2002; Cordin *et al.*, 2006). RecA-like domain refers to a “core” α/β structure module of a large central parallel β -sheet flanked by α -helices, homologues in structure to the core region of RecA protein (Story and Steitz, 1992). The two RecA-like domains are in parallel with the NTPase active site at the interface and the NA binding motifs are on the top of the two domains (Caruthers and McKay, 2002). Additional domain insertions are commonly seen in different helicases (Caruthers and McKay, 2002; Jankowsky and Fairman, 2007).

Figure 1. 20 Topology diagrams of RecA and RecA-like core of the SF1 and SF2 helicases. Yellow, conserved RecA-related ‘core’. Red, variable structural elements. Green, additional structural domain. The schematic on the right summarizes the

positions in the topology of the RecA-like core of the seven conserved motifs defined by (Gorbalenya and Koonin, 1993) for the SF1 and SF2 helicases. Adapted from Caruthers and MaKay, 2002.

1.3.3.1.1 Upf1 helicase core (SF1)

SF1 DNA helicases like PcrA, Rep, and UvrD share similar domain arrangements. The two RecA-like domains are named 1a and 2a on top of which are the two additional domains, 1b and 2b, which predominantly consist of α -helices (Korolev *et al.*, 1997; Lee and Yang, 2006; Velankar *et al.*, 1999). Additional structural motifs involved in ssDNA or dsDNA binding are identified from domain 1b and 2b (Lee and Yang, 2006; Velankar *et al.*, 1999). Upf1 helicase core is the only RNA helicase belonging to SF1 with known atomic structure (Cheng *et al.*, 2007b). The core RecA-like domain arrangement (1a and 2a) is found in common with other SF1 and SF2 helicases. However, unique domain insertions (1b and 1c) are identified, that are not closely related to any of the known helicase structures. Domain 1b, is a β barrel structure composed of 6 anti parallel β strands. It is located above the interface between domains 1a and 2a. Domain 1c contains 3 α helices and is in contact with both domain 1a and 1b. A putative ssRNA binding groove is formed by domain 1a, 2a and 1b well aligned to that in PcrA or Vasa. A channel between domain 1b and 1c might also be part of the ssRNA binding groove accommodating the 3' portion of the NA. Domain 1b appears to modulate the coupling of ATP binding to and RNA binding and domain 1c is required for RNA binding and maximal ATP hydrolysis rates *in vitro*. It was suggested protein structural rearrangements induced by ATP binding might inhibit Upf1 helicase – ssRNA interaction, while ADP binding might enhance RNA binding (Cheng *et al.*, 2007b). However, the exact structural and mechanical detail is not clear. Furthermore, how ATP hydrolysis and RNA binding by Upf1 is related to its function in nonsense-mediated mRNA decay (NMD, one mRNA surveillance pathway) is still elusive.

1.3.3.1.2 DE-A/x-D/H RNA helicases (SF2)

Most RNA helicases belong to SF2 and are grouped into DEAD box, DEAH and DExH subfamilies (Jankowsky and Fairman, 2007). Many crystal structures of SF2 RNA helicases are available either as apo enzyme or in complex with nucleotides and/or NA. **Table 1. 2** summarizes the structures of RNA helicases from SF1/SF2 available from PDB. The structures containing only one of the two RecA-like domains are not included in the table nor in our discussion.

1.3.3.1.3 DExH helicase: HCV NS3

The C-terminal portion of HCV NS3 is the first and the most intensively studied SF2 RNA helicase (Frick, 2007). Fueled by ATP hydrolysis, HCV NS3 can translocate along ssRNA or ssDNA in a 3' to 5' direction, unwind NA duplex (dsDNA/dsRNA) and displace proteins bound to the NA (Kim *et al.*, 1995; Suzich *et al.*, 1993; Tai *et al.*, 1996). In contrast to the well known SF1 helicases like Upf1, PcrA, Rep, and UvrD, HCV NS3 helicase only has three domains, with two parallel RecA-like cores and an additional domain at the C terminus that is composed mainly of α helices (Cho *et al.*, 1998; Kim *et al.*, 1998; Mackintosh *et al.*, 2006; Yao *et al.*, 1997; Yao *et al.*, 1999). The seven motifs of SF2 helicases are located in domains 1 and 2 next to the interdomain clefts. The structure reported by (Kim *et al.*, 1998) is in complex with a ssDNA oligonucleotide (dU₈). The ssDNA lies in a groove between the two RecA-like domains and domain 3 with the 5' end of the ssDNA residing at the interface of domains 2 and 3 and its 3' end at the interface of domains 1 and 3. Interactions between the ssDNA and enzyme are mostly confined to the DNA backbone, as would be expected for a nonspecific protein-NA complex. In particular, Trp-501 stacks with the base of dU₈ and Val-432 interacts with the dU₄ base. These two side chains act as a pair of "bookends", defining a central binding cavity occupied by five nucleotides (Kim *et al.*, 1998). The crystal structure reported recently by (Mackintosh *et al.*, 2006) contains three HCV NS3 helicase monomers per asymmetric unit and in complex with an ssDNA long enough to accommodate binding of two monomers. In general, both proteins bind to the ssDNA in the same manner as that reported earlier (Kim *et al.*, 1998). Several amino acid residues at the interface of the two HCV NS3 molecules were identified to mediate a protein-protein

interaction between domains 2 and 3 of adjacent molecules, suggesting NS3 may work as dimer (Mackintosh *et al.*, 2006). Binding of an oligonucleotide did not induce large scale structural rearrangement to the protein, although open and closed forms of apo HCV NS3 are reported separately (Cho *et al.*, 1998; Yao *et al.*, 1997). Unfortunately, there is no structure available with a bound nucleotide either mimicking the substrate binding state or the product binding state. The mechanism of ATP hydrolysis is speculated by applying structural modeling and mutagenesis data (Frick *et al.*, 2007).

In the crystal structure of full length NS3-NS4A, the protease domain packs behind the helicase domain and receives its cofactor NS4A, which extends from the C-terminus of helicase domain to its catalytic site (Yao *et al.*, 1999). Interestingly, the protease domain and the helicase domain enhance each other's enzymatic activity (Beran and Pyle, 2008; Beran *et al.*, 2007).

1.3.3.1.4 DEAH helicases

The structure of DENV2 NS3 helicase domain is introduced in Section 1.2.4.4.2. The other Flavivirus NS3 helicases have almost identical structure to DENV2 NS3 helicase, so are the helicase motifs (Mancini *et al.*, 2007; Mastrangelo *et al.*, 2007c; Speroni *et al.*, 2008; Wu *et al.*, 2005; Xu *et al.*, 2005; Yamashita *et al.*, 2008).

1.3.3.1.5 DEAD box helicases

DEAD box helicases are the largest family of RNA helicases by far and the different DEAD box helicases usually play highly specific roles in RNA metabolism. For instance, most of the 26 DEAD box helicases present in yeast are essential, indicating that they cannot replace each other (Bleichert and Baserga, 2007). Specificity would arise from the fact that DEAD box helicases generally do not work in isolation but in association with specific partners that modulate their activity and direct them to specific targets (Silverman *et al.*, 2003). The N-terminal or C-terminal extensions that flank the helicase core are thought to be mainly responsible for these associations (Rocak and Linder, 2004).

Although the function of the DEAD box proteins in mRNA metabolism is different, the structurally conserved core element of the DEAD box helicases and the way they bind RNA and ATP is essentially the same. In general, DEAD box helicases contain two conserved RecA-like domains that adopt different open structures in the absence of NA (Caruthers *et al.*, 2000; Cheng *et al.*, 2005; Shi *et al.*, 2004; Story *et al.*, 2001) and a closed structure when bound to RNA and ATP (Andersen *et al.*, 2006; Bono *et al.*, 2006; Nielsen *et al.*, 2009; von Moeller *et al.*, 2009).

Table 1. 2 Structures of RNA helicases from SF1/SF2

Group	Protein	PDB ID	Remarks (Ligands / Cofactors)	Reference
SF1	Upf ^H	2GJK	Mg ²⁺ / AMPPNP	(Cheng <i>et al.</i> , 2007b)
		2GK6	Mg ²⁺ / ADP	
		2GK7	PO ₄	
SF2 (DEAD)	eIF4A ^Y	1FUU	-	(Caruthers <i>et al.</i> , 2000)
		2VSO ^{P1}	AMP / eIF4G	(Schutz <i>et al.</i> , 2008)
		2VSX ^{P21}	AMP / eIF4G	
	eIF4A ^H	2ZU6	PDCD4	(Chang <i>et al.</i> , 2009)
		3EIQ	PDCD4	(Loh <i>et al.</i> , 2009)
	mjDEAD ^A	1HV8	SO ₄	(Story <i>et al.</i> , 2001)
	UAP56 ^H	1XTI	-	(Shi <i>et al.</i> , 2004)
		1XTJ	ADP / Mg ²⁺	
		1XTK	DECD to DEAD mutation	
	Dhh1p ^Y	1S2M	-	(Cheng <i>et al.</i> , 2005)
	Vasa ^D	2DB3	AMPPNP / Mg ²⁺ / ssRNA	(Sengoku <i>et al.</i> , 2006)
		2J0S	AMPPNP / Mg ²⁺ / ssRNA (EJC)	(Bono <i>et al.</i> , 2006)
	eIF4AIII ^H	2J0U	Barentsz	
		2HXY	-	(Andersen <i>et al.</i> , 2006)
		2HYI	AMPPNP / Mg ²⁺ / ssRNA (EJC)	
		3EX7	ADP / AIF ₃ / Mg ²⁺ / ssRNA (EJC)	
	DDX3X ^H	2I4I	AMP	(Hogbom <i>et al.</i> , 2007)
	stDEAD ^A	2Z0M	-	(Nakagawa <i>et al.</i> , 2007)
	DDX19 ^H	3EWS	ADP	(Lehtio <i>et al.</i> , 2008)
		3G0H	AMPPNP / Mg ²⁺ / ssRNA	(Karlberg <i>et al.</i> , 2009)
	Dbp5 ^Y	3FHT	AMPPNP / Mg ²⁺ / ssRNA	(von Moeller <i>et al.</i> , 2009)
SF2 (DExH)	HCV NS3	1HEI	SO ₄	(Yao <i>et al.</i> , 1997)
		1A1V	SO ₄ / ssDNA	(Kim <i>et al.</i> , 1998)
		8OHM	-	(Cho <i>et al.</i> , 1998)
		1CU1	PO ₄ / NS3-NS4A Protease	(Yao <i>et al.</i> , 1999)
		2F55	SO ₄ / ssDNA	(Mackintosh <i>et al.</i> , 2006)
SF2 (DEAH)	YFV NS3	1YKS	-	(Wu <i>et al.</i> , 2005)
		1YMF	ADP	
	DENV2 NS3	2BMF	SO ₄	(Xu <i>et al.</i> , 2005)
		-	Mn ²⁺ / SO ₄	(Xu, 2007)
	MVE NS3	2V8O	-	(Mancini <i>et al.</i> , 2007)
	WNV NS3	2QEQ	-	(Mastrangelo <i>et al.</i> , 2007c)
	Kokobera NS3	2V6I	pyrophosphate	(Speroni <i>et al.</i> , 2008)
	JEV NS3	2Z83	-	(Yamashita <i>et al.</i> , 2008)

^A, Archaea; ^D, Drosophila; ^H, Human; ^Y, Yeast; EJC, Exon Junction Complex.

1.3.3.2 Mechanism of RNA Helicases

In general, unwinding is viewed as a combination of (1) RNA binding, (2) unidirectional translocation and (3) strand separation processes. Both of these processes are fueled by the energy released from NTP hydrolysis (Caruthers and McKay, 2002; Pyle, 2008). Based on this concept, models of helicase translocating along single stranded and/or unwinding duplex NA are proposed and briefly discussed below. Each model is apparently supported by certain *in vitro* experiments on certain helicases. The dynamic process would require at least alternations of tight and loose binding of the tracking strand RNA and coupling to the substrate nucleotide binding, hydrolysis and product release. Apparently, there is no single model could explain all phenomena, regardless of variations between different measurement techniques and analysis methodologies. Furthermore, the different helicases might simply evolve to have different molecular mechanisms to perform different molecular tasks in different environments.

1.3.3.2.1 Theoretical model: passive vs active

A passive helicase waits for a thermal fluctuation that opens part of the dsNA, and then moves forward and binds to the newly available single strand. Passive opening corresponds to a hard-wall potential: the helicase inhibits NA closing, but does not otherwise affect the duplex. The base pair adjacent to the helicase has a probability α/β of being open. Thus, when the helicase attempts a forward hop it succeeds with probability α/β (**Figure 1. 21**).

An active helicase directly destabilizes the duplex. A softer staircase potential corresponds to active opening. To accelerate unwinding, it is better to increase α than to decrease β because faster opening increases the strand separation time scale. The interaction potential can be chosen to maximize the unwinding rate, and it is calculated that an optimized active helicase can unwind duplex as fast as it translocates on ssNA (**Figure 1. 21**).

Changes in the ATP concentration might affect the hopping rates k^+ and k^- , leading to changes in the unwinding rate. Similarly, external forces or interaction with extra NA binding domain or proteins can affect the free energy of melting duplex (α/β). The validity of this theoretic model is yet to be verified by accurately

measuring both translocation, unwinding rates and the consumption of energy (ATP hydrolysis) of different helicases (Betterton and Julicher, 2005).

Figure 1. 21 Sketch of helicase on NA. The helicase moves forward toward the duplex at rate k^+ and backward at rate k^- . The NA opens at rate α and closes at rate β . The duplex opens and closes due to thermal fluctuations. Thermal breathing is defined by the rates α and β at which the NA base pair at the ss/ds junction opens and closes, respectively. For simplicity, it is assumed that these rates are independent of the NA base sequence. Adapted from Betterton and Julicher, 2005.

1.3.3.2.2 Active rolling model

The active rolling model was originally proposed based on the studies of SF1 DNA helicase Rep from *Escherichia coli* (Figure 1. 22)(Korolev *et al.*, 1997). In the active rolling model, the dimeric helicase interacts with both dsDNA and ssDNA. Each subunit alternates binding to dsDNA as the dimer translocates when one subunit releases ssDNA and rebinds to dsDNA. Translocation along ssDNA is coupled to ATP binding (Tuteja and Tuteja, 2004). Many helicases are proven active as monomers so that in these cases this model is not applicable (Cheng *et al.*, 2007a; Dessinges *et al.*, 2004; Niedziela-Majka *et al.*, 2007).

Figure 1. 22 Active rolling model for DNA unwinding and translocation. The

letter (P) indicates the Rep monomer. The letters (S) and (D) indicate single-stranded DNA and double-stranded DNA, respectively. Adapted from Korolev *et al.*, 1997.

1.3.3.2.3 Inchworm model

In the inchworm model, the enzyme monomer is bound to ssDNA and then translocates along the DNA strand to the fork region, probably upon binding ATP (**Figure 1. 23**). Helix destabilization and release of one of the ssDNA strands takes place as ATP is hydrolyzed (Velankar *et al.*, 1999). The ligand-induced conformation changes on PcrA bound to the same DNA substrate provide evidences consistent with an “inchworm” mechanism (Velankar *et al.*, 1999). However, the mechanism of duplex opening is not explained by this model. The additional interactions between domain 2a with ss/ds fork and domain 2b with the upstream dsDNA were proposed to trigger duplex melting at the ss/ds fork (Lee and Yang, 2006; Velankar *et al.*, 1999).

Figure 1. 23 Inchworm model. A typical helicase monomer contains two RecA-like domains in a tight (closed hand) and a weak (open hand) nucleic acid state. The protein undergoes steps of helicase movement (protein conformational changes) and nucleic acid affinity changes (tight to weak transitions) upon ATP binding and hydrolysis. Adapted from Velankar *et al.*, 1999.

1.3.3.2.4 “Brownian motor” model

A “Brownian motor” exploits random movements that constantly occur on the molecular level (Brownian motion) and an asymmetrical path to shift an object in a single direction. The key to the Brownian motor model is the asymmetry of the path on which the motor is traveling so that the net result is unidirectional movement. Based on the kinetic studies on interactions of the HCV NS3 helicase with various DNA substrates, Levin *et al.* proposed a Brownian motor mechanism for the

directional movement and strand separation by the helicase (**Figure 1. 24**) (Levin *et al.*, 2005).

Figure 1. 24 Brownian motor mechanism of translocation and unwinding. (A) The binding free energy (ΔG) of the helicase-ssDNA complex (gray circles) changes along the length of the DNA (solid lines) and the sawtooth profile is shaped by specific interactions in the complex. In the absence of ATP, the helicase binds ssDNA tightly and settles into a deep energy minimum unlikely to move (position 1). ATP binding changes the properties of the DNA binding site, which becomes weaker and almost constant along DNA (position 2). In this transient state, the helicase moves along the DNA randomly and bidirectionally. Rapid hydrolysis of the bound ATP restores the asymmetric sawtooth profile of the complex and forces the helicase to slide down the energy slope; that is, to make a power stroke. Depending on the direction of the net random movement during the previous stage, the helicase lands either in the same minimum (position 1) or one step ahead (position 3). Binding of the new ATP molecule allows the helicase to start the next translocation cycle (position 4). (B) A mechanism that produces the energy profile needed for translocation. Adapted from Levin *et al.*, 2005.

The unwinding mechanism tends to be passive, depending on ss/ds junction opening by thermal fluctuations (Levin *et al.*, 2005). Multimerization might be a mean to improve the unwinding efficiency (Ali *et al.*, 1999; Jennings *et al.*, 2008; Levin *et al.*, 2004).

1.3.3.2.5. Local strand separation by DEAD box helicases

A recent study on RNA unwinding by several DEAD box helicases suggests a new model of unwinding in contrast to the unwinding mechanisms of previously characterized DNA and RNA helicases (**Figure 1. 25**) (Yang *et al.*, 2007). The model explains the distinct characteristics of the DEAD box helicases, poor processivity, duplex unwinding without preferred polarity, the ability to unwind blunt-end substrates, and strand annealing activities. (Bleichert and Baserga, 2007; Jankowsky and Fairman, 2007; Rajkowitsch *et al.*, 2007). It is also consistent with the physiological functions of the DEAD box helicases as RNA chaperones for remodeling of RNA/RNP structures (Bleichert and Baserga, 2007; Cordin *et al.*, 2006; Rocak and Linder, 2004).

Figure 1. 25 Three-step mechanism for duplex unwinding by DEAD box helicases. Step 1: The DEAD box helicase binds to the duplex either at a terminal or internal region. This loading is facilitated by ssRNA region. Step 2: The duplex, wherever the RNA residues are accessible, is opened locally in an ATP-dependent reaction. Step 3: Dissociation of remaining base pairs. For longer and/or more stable

duplexes, this step will greatly affect the overall efficiency of the strand separation. Determinants of the overall duplex unwinding rate are the loading efficiency (step 1, accounting for effects of single-stranded regions and effects of enzyme concentrations), the rate by which the duplex is locally opened (step 2, accounting for effects of ATP concentration), and the lifetime of this locally opened state, as well as on the stability of remaining base pairs (step 3, accounting for the effect of duplex length and/or stability). Adapted from Yang *et al.*, 2007.

1.4. Aims of the project

Dengue fever is an emerging public health concern, with several million viral infections occurring annually. No effective therapy currently exists. NS3, the second largest viral protein, is endowed with viral protease, helicase, nucleoside 5'-triphosphatase and 5'-terminal RNA triphosphatase activities. In the recent years, several structures of NS3 protease and helicase individual domains were determined and insights about their individual enzymatic activities were gained from the structures.

As a multi-functional enzyme, NS3 works together with NS2B for viral polyprotein processing and with NS5 as part of the viral replication complex. Hence, structural studies of the full length NS3 polypeptide are needed to understand the molecular basis for its dynamic properties in connection to its various enzymatic activities and to gain structural insights about the dynamic assemblies of the machineries for viral polyprotein processing and viral replication.

NS3 helicase is an SF2 RNA helicase, sharing a broad range of similarities with other RNA or DNA helicases. However, there are structural and functional differences among different groups of helicases. Moreover, our understanding about the detailed mechanisms for ATP hydrolysis, ssRNA binding and dsRNA unwinding activities of the enzyme is still very limited. Questions about how the chemical energy is utilized for mechanical work and about helicase properties like unwinding speed, step size, processivity and substrate preferences remain to be answered. Consequently structural studies of NS3 helicase with its ligands (nucleotide and nucleic acids) will provide useful information and help address these questions.

1.5 Presentation of the work

Previous research on flavivirus NS3 protein was reviewed before the initiation of this work: NS3 protease domain (**NS3Pro**) is not very soluble on its own (Murthy *et al.*, 1999); Full length NS3 (**NS3FL**) is not very soluble either (Sampath *et al.*, 2006; Xu *et al.*, 2005); The central hydrophilic region of NS2B is essential for NS3 protease activity, the presence of this cofactor peptide (**NS2B₄₇**, 47 amino acids) lead to very soluble and active protease, and crystallographic structures of this NS2B₄₇NS3Pro were determined (Erbel *et al.*, 2006); Furthermore, the presence of the first 18 amino acids from the NS2B cofactor region (**NS2B₁₈**) renders NS3 protease soluble but enzymatically inactive (Erbel *et al.*, 2006); Based on the structures of NS2B₄₇NS3Pro, NS2B₁₈NS3Pro shall retain the same overall protein fold. Autolysis occurs to the NS2B₄₇NS3Pro and NS2B₄₇NS3FL proteins (Bera *et al.*, 2007; Erbel *et al.*, 2006).

Protein solubility and stability are the key properties that are necessary in order to perform structural studies on NS3FL. Based on the observations mentioned above, NS2B₁₈NS3FL construct was made (**Figure 3.1A**). By including NS2B₁₈, soluble NS3FL protein could be produced to large quantity; NS2B₁₈NS3 shall not cleave itself as its protease domain is inactive. Thus by linking NS2B₁₈ and NS3FL with a flexible nano-peptide linker (Gly₄-Ser-Gly₄), a single chain NS3 protease-helicase protein (**scNS2B₁₈NS3**) was engineered for structural and functional studies in the first part of this work (**Section 3.1**). The C terminal helicase domain construct (**NS3hel**, amino acids 172-618) was made later on based on the protein stability analysis of scNS2B₁₈NS3 protein (**Figure 3.1C**). This NS3hel protein was used for structural and functional studies in the second part of this work (**Section 3.2**).

Chapter 2 Materials and Methods

2.1 Molecular cloning

scNS2B₁₈NS3. DENV4 viral RNA (Source: NITD dengue virus sample storage) was prepared using the QIAamp viral RNA Mini Kit (Qiagen). cDNA was generated using Reverse Transcriptase PCR (Invitrogen SuperScriptTM III) with the full-length NS3 reverse primer (5'-GTGGTGCTCGAGTTACTTTCTTCCACTGG CAAA-3'). The underlined sequence corresponds to the *XhoI* site. The fragment corresponding to amino acid residues 1 to 618 of NS3 was amplified from viral cDNA using the forward primer (5'-GGGGGCGGAGGTAGTGGTGGAGGCGGT CAGGAGCTCTGTGGGACGTC-3') and the reverse primer above. The nucleotide sequence encoding the Gly₄SerGly₄ linker sequence is in italics. The central hydrophilic fragment corresponding to amino acid residues 49 to 66 of the NS2B protein (NS2B₁₈) was amplified from viral cDNA using the forward (5'-GCAGACTTGTCCTAGAGAAG-3') and reverse primers (5'-CCCGCCTCCA CCACTACCTCCGCCCGTCCGCCATTTCATCCCA-3') with the linker sequence in italics. A third PCR reaction was subsequently performed to generate the complete NS2B₁₈ Gly₄SerGly₄ NS3 fusion construct using the NS2B₁₈ forward and NS3 reverse primers listed above. The fragment was digested with *XhoI* and cloned into a modified pET32b plasmid using T4 ligase (Roche), where the S-tag and enterokinase cleavage sequence are absent. The expressed construct thus comprises the Thioredoxin (Trx) protein followed by a hexa-Histidine tag and a Thrombin cleavage site fused at the N-terminus of scNS2B₁₈NS3.

NS3_{hel} (residues 172-618) was amplified by PCR from the full length DENV4 NS3 gene using forward primer (5'-ATTGGTGAGCCAGATTA TGAAGTGG-3') and the reverse primer (5'-GTGGTGCTCGAGTTACTTCTTCCACTGGCAAA-3'). The *XhoI* site is underlined. The fragment was digested and cloned into a modified pET32b plasmid using T4 ligase (Roche), where the S-tag and enterokinase cleavage sequence are absent. The expressed construct thus comprises the Thioredoxin (Trx) protein followed by a hexa-Histidine tag and a Thrombin cleavage site fused to the N-terminus of NS3h.

NS3hel (residues 177-618) was cloned by the same method as DENV4 NS3hel (residues 172-618), using the forward primer (5'-GAAGTGGATGAG GACATT-3') and reverse primer (5'-GTGGTGCTCGAGTTACTTTCTTCCACT GGCAAA -3'), with the *XhoI* site underlined.

2.2 Protein expression and purification

scNS2B₁₈NS3. (A) Recombinant protein prepared for crystallization of Conformation I (PDB code: 2BVC). Transformed *E. coli* BL21-CodonPlus clones (Stratagene) were grown at 18 °C in an auto induction medium (Studier, 2005) supplemented with 100 µg ml⁻¹ ampicillin and 50 µg ml⁻¹ chloramphenicol. After 28 hrs, cells were harvested by centrifugation at 8,000 × g for 10 min at 4°C and stored at -20°C. Cells resuspended in buffer A (20 mM Na₃PO₄, pH 7.4, 0.5 M NaCl, 40 mM imidazole) were lysed by sonication and the lysate clarified by centrifugation at 30,000 × g for 60 mins at 4°C. The supernatant was purified by metal affinity using a HisTrap HP column (Amersham Bioscience) equilibrated with buffer A. Proteins were eluted using a linear gradient of imidazole from 40 mM to 500 mM. The fraction containing Trx-(His)₆- scNS2B₁₈NS3 was dialyzed against modified PBS buffer (200 mM NaCl, 10 mM Phosphate, 2.7 mM KCl, and a pH of 7.4), with concomitant cleavage of the thioredoxin tag by thrombin digestion (substrate/enzyme ratio of 500:1) at 4°C for approximately 24 h. The cleavage mixture was loaded onto a HisTrap HP column equilibrated with buffer B (20 mM Na₃PO₄, pH 7.4, 0.2 M NaCl) in order to remove the Trx-(His)₆ protein from the mixture. Proteins were further purified using anion-exchange chromatography (MonoQ HR 5/5, Amersham Bioscience) and eluted using a linear gradient ranging from 0.05 to 1.0 M NaCl in buffer C (20 mM Tris-Cl, pH 8.0, 0.05 M NaCl). Concentrated scNS2B₁₈NS3 proteins were subjected to a final polishing step through a HiPrep Superdex-200 gel-filtration column (Amersham Bioscience) in buffer D (20 mM Tris-HCl, pH 7.4, 0.2 M NaCl, 1 mM dithiothreitol, 5% Glycerol). Fractions containing scNS2B₁₈NS3 were pooled and concentrated to 20 mg ml⁻¹. Protein concentration was determined by measuring the absorbance of the protein solution at 280 nm (1 unit A_{280nm} corresponds to 0.7 mg ml⁻¹ based on the calculated protein molecular weight of 72.3 kDa and molecular extinction coefficient of 103,990 M⁻¹ cm⁻¹). (B) Recombinant protein prepared for the crystallization of

Conformation II (PDB code: 2WHX) and other purposes. Transformed *E. coli* BL21-CodonPlus cells (Stratagene) were grown at 37 °C in LB medium supplemented with 100 µg ml⁻¹ ampicillin and 50 µg ml⁻¹ chloramphenicol to an OD_{600nm} of 0.6 to 0.8. Protein expression was induced at 16°C by adding isopropyl-β-D-thiogalactopyranoside (IPTG) with a final concentration of 0.4 mM. After overnight growth, cells were harvested by centrifugation at 8,000 × g for 10 min at 4°C and stored at -20°C. Cells resuspended in buffer A were lysed by sonication and the lysate clarified by centrifugation at 30,000 × g for 60 minutes at 4°C. The supernatant was purified by metal affinity using a HisTrap HP column (Amersham Bioscience) equilibrated with buffer A. Proteins were eluted using a linear gradient of imidazole from 40 mM to 500 mM. The fraction containing Trx-(His)₆- D4NS3hel was dialyzed against modified PBS buffer (200 mM NaCl, 10 mM Phosphate, 2.7 mM KCl, and a pH of 7.4), with concomitant cleavage of the Thioredoxin tag by thrombin (10 enzyme units per mg of protein) at 4°C for approximately 24 h. The cleavage mixture was loaded onto a HisTrap HP column equilibrated with buffer B in order to remove the Trx-(His)₆ protein from the mixture. Concentrated scNS2B₁₈NS3 proteins were subjected to a final step through a HiPrep Superdex-200 gel-filtration column (Amersham Bioscience) in buffer D' (20 mM Tris-HCl, pH 7.4, 0.15 M NaCl, 5 mM dithiothreitol, 5% Glycerol). Fractions containing scNS2B₁₈NS3 were pooled and concentrated to 20 mg ml⁻¹ or above with batch to batch variations.

NS3hel (residues 172-618). Transformed *E. coli* BL21-CodonPlus cells (Stratagene) were grown at 37 °C in LB medium supplemented with 100 µg ml⁻¹ ampicillin and 50 µg ml⁻¹ chloramphenicol to an OD_{600nm} of 0.6 to 0.8. Protein expression was induced at 16°C by adding isopropyl-β-D-thiogalactopyranoside (IPTG) with a final concentration of 0.4 mM. After overnight growth, cells were harvested by centrifugation at 8,000 × g for 10 min at 4°C and stored at -20°C. Cells resuspended in buffer A were lysed by sonication and the lysate clarified by centrifugation at 30,000 × g for 60 minutes at 4°C. The supernatant was purified by metal affinity using a HisTrap HP column (Amersham Bioscience) equilibrated with buffer A. Proteins were eluted using a linear gradient of imidazole from 40 mM to 500 mM. The fraction containing Trx-(His)₆- NS3hel was dialyzed against modified PBS buffer (200 mM NaCl, 10 mM Phosphate, 2.7 mM KCl, and a pH of 7.4), with concomitant cleavage of the thioredoxin tag by thrombin (10 enzyme units per mg of

protein) at 4°C for approximately 24 h. The cleavage mixture was loaded onto a HisTrap HP column equilibrated with buffer B in order to remove the Trx-(His)₆ protein from the mixture. Concentrated NS3hel proteins were subjected to a final polishing step through a HiPrep Superdex-75 gel-filtration column (Amersham Bioscience) in buffer D' (20 mM Tris-HCl, pH 7.4, 0.15 M NaCl, 5 mM dithiothreitol, 5% Glycerol). Fractions containing pure NS3hel protein were pooled and concentrated to 20 mg ml⁻¹ or above with batch to batch variations. Protein concentration was determined by measuring the absorbance of the protein solution at 280 nm (1 A_{280nm} unit corresponds to 0.78 mg ml⁻¹ based on the calculated protein molecular weight of 51.4 kDa and molecular extinction coefficient of 65,890 M⁻¹ cm⁻¹). DENV4 NS3hel (residues 177-618) protein was expressed and purified by Dr Xuting, following the same protocol. The protein was concentrated to 34 mg ml⁻¹ and stored in buffer D'' (20 mM Tris-HCl, pH 7.4, 50mM NaCl, 5 mM β-mercaptoethanol, 5% Glycerol).

2.3 Matrix-assisted laser desorption ionization time-of-flight mass spectrometry.

The integrity of scNS2B₁₈NS3 protein was checked by MALDI-TOF, API 300 MS/MS; Applied Biosystems (Proteomic Core Facility, School of Biological Sciences, Nanyang Technological University). Results for the freshly prepared scNS2B₁₈NS3 and the proteolysis sample after 168 hours incubation at RT are shown in **Figure 3.1 A/C**.

2.4 N-terminal amino-acid sequence determination.

Automated N terminal amino acid sequence determination of the proteolytic fragments derived from scNS2B₁₈NS3 was performed using an Applied Biosystems Procise sequencer (Protein Proteomics Center, National University of Singapore, Singapore). The protein sample was pre-incubated 168 hours at RT (equivalent to Lane 6 in **Figure 3.1C**).

2.5 Crystallization and data collection

scNS2B₁₈NS3. (A) Conformation I (PDB code: 2BVC). Crystals were grown at 18°C by the hanging drop vapor diffusion method over wells containing 0.1 M morpholino-ethanesulfonic acid (MES) pH 6.5, 7.5% polyethylene glycol 3,350. A volume of 2 μ l of precipitating solution was mixed with an equal volume of scNS2B₁₈NS3 at a concentration of 3-5 mg ml⁻¹. Crystals grew as clusters of thin elongated plates over 2-5 days to dimensions of approximately 0.02 $\times$ 0.30 $\times$ 0.10 mm³ (**Figure 2. 1**). For data collection, crystals were soaked in a solution containing (0.1 M MES, 20% PEG 3,350 at pH 6.5) for 30 minutes at room temperature and were subsequently soaked for 1 minute in (25% glycerol, 0.1 M MES, 20% PEG 3,350 at pH 6.5), before being mounted and cooled to 100 K in a nitrogen gas stream (Oxford cryosystems). Diffraction intensities were recorded on an ADSC charge-coupled device detector on the ID-23-1 beamline at the European Synchrotron Radiation Facility (ESRF, Grenoble, France), using an attenuated beam of dimensions 0.100 $\times$ 0.100 mm². Integration, scaling and merging of the intensities were carried out using programs MOSFLM and SCALA from the CCP4 suite (Collaborative, 1994). The crystal parameters and data collection statistics are summarized in (B) Conformation II (PDB code: 2WHX). Crystals of Conformation II were grown at 13°C by the hanging drop vapor diffusion method over wells containing 0.1 M morpholino-ethanesulfonic acid (MES) pH 6.5, 5% polyethylene glycol 3,350. A volume of 2 μ l of precipitating solution was mixed with an equal volume of scNS2B₁₈NS3 at a concentration of 3-5 mg ml⁻¹. Crystals grew as clusters of thin elongated plates over 2-5 days, similar to that for Conformation I. For data collection, crystals were soaked for 20 hours in (25% glycerol, 0.1 M MES, 25% PEG 3,350 at pH 6.5, with 5 mM MnCl₂ and 5 mM ADP), before being mounted and cooled to 100 K in a nitrogen gas stream (Oxford cryosystems). Diffraction intensities were collected at NSRRC, Beamline 13B1 (Hsinchu, Taiwan, ROC). Integration, scaling and merging of the intensities were carried out using programs MOSFLM and SCALA from the CCP4 suite (CCP4, 1994). The crystal parameters and data collection statistics are summarized in table 2.1.

(A)

(B)

Figure 2. 1 scNS2B₁₈NS3 crystals. (A) Typical scNS2B₁₈NS3 crystals are thin plates forming clusters. (B) SDSPAGE confirms the protein molecules that are packed in the crystals are the intact scNS2B₁₈NS3 protein. It also indicates that protein molecules that are packed in the crystal lattice are resistant to proteolysis.

NS3hel (residues 172-618). Crystals of native NS3hel (**Figure 2. 2A**) (storage buffer: buffer D') were grown at 13 °C by the hanging drop vapor diffusion method

over wells containing 0.1 M 2-(*N*-morpholino)ethanesulfonic acid (MES), pH 6.5, 20% polyethylene glycol 3,350. A volume of 2 μ l of precipitating solution was mixed with an equal volume of NS3hel at a concentration of 5 mg ml⁻¹. The drop was equilibrated against a reservoir containing 1 ml of the precipitating solution. Crystals for the AMPPNP (Roche) complex were obtained by co crystallization of NS3hel (storage buffer: buffer D') at a concentration of 5 mg ml⁻¹ (~0.1 mM), with 5 mM MnCl₂ and 5 mM AMPPNP using a slightly different precipitating solution (0.1 M MES, pH 6.5, 10% polyethylene glycol 3,350) at 13 °C. Crystals with ADP (Roche) were obtained by co-crystallization (storage buffer: buffer D'' (20 mM Tris-HCl, pH 7.4, 50mM NaCl, 5 mM β -mercaptoethanol, 5% Glycerol)) at a concentration of 2.5 mg ml⁻¹ with 5 mM MnCl₂ and 5 mM ADP in the precipitating buffer (0.1 M Tris-HCl, pH 7.0, 7.5% polyethylene glycol 3,350) at 23 °C. The binary complex with RNA₁₂ was also obtained through co crystallization: a final solution containing NS3hel (storage buffer: buffer D') at 2.5 mg ml⁻¹ and RNA₁₂ (Sigma) at 0.1 mM (~ 2 fold molar excess) was incubated at 25°C for 1 hour. Subsequently, a volume of 3 μ l of the NS3hel-RNA₁₂ solution was mixed with 2 μ l of precipitating solution (0.1 M MES, pH 6.0, 15% polyethylene glycol 8,000) and the drop was equilibrated against a reservoir containing 1 ml of the precipitating solution (**Figure 2. 2B**). Interestingly, using identical co-crystallization conditions, isomorphous crystals can be obtained either with a AGU₁₀ 12-mer or with the 5' untranslated region 5'UTR (AGUUGUUAGUCU) of the viral genome, but not with U₁₂ suggesting that a Guanine at position 2 is needed for crystallization with a 12-mer.

The ternary complexes containing RNA₁₂ and either AMPPNP, ADP-P_i, ADP/Vanadate or ADP respectively (**Table 2. 3**) were obtained through 20 hours soak of crystals of the binary RNA complex in a cryoprotecting solution (0.1M MES, pH 6.0, 25% PEG8,000, 25% glycerol) that contained 5 mM MnCl₂ in the presence of either 5mM AMPPNP, ATP, ADP and orthovanadate (Sigma) or ADP at 13 °C. Prior to data collection, native and AMPPNP crystals were briefly soaked in a cryoprotecting solution containing 0.1M MES, pH 6.5, 20% PEG3,350, 25% glycerol; ADP complex crystals in 0.1M Tris-HCl, pH 7.0, 20% PEG3,350, 25% glycerol; RNA-bound crystals in 0.1 M MES, pH 6.0, 25% PEG 8,000, 25% glycerol for 20 hours. Following cryoprotection, crystals were rapidly cooled to 100 K in a nitrogen gas stream (Oxford cryosystems) and diffraction intensities were recorded at the

Chapter 2 Materials and Methods

Nanyang Technological University (Singapore) on an R-AXIS IV++ imaging plate detector. Integration, scaling and merging of the intensities were carried out using programs of MOSFLM and Scala (CCP4, 1994). For the RNA₁₂/ADP/VO₄/Mn²⁺ and RNA₁₂/AMPPNP/ Mn²⁺ ternary complexes data at higher resolution were collected at the ESRF, Grenoble (beamline ID23-1) and at the Swiss Light Source, Villingen (beamline X10SA) respectively, using charged-coupled-device detectors.

(A)

(B)

Figure 2. 2 DENV4 NS3hel crystals. (A) Crystals for NS3hel apo enzyme. (B) NS3hel-ssRNA crystals.

2.6 Structure solution and refinement

scNS2B₁₈NS3. (A) Conformation I (PDB code: 2VBC). The crystal contains one monomer per asymmetric unit with an estimated solvent content of 50.2 % based on a V_m value of 2.47 (Matthews, 1968). The structure was solved by molecular replacement with the program PHASER (Storoni *et al.*, 2004), using the DENV2 NS3pro (PDB code: 2FOM) (Erbel *et al.*, 2006) and DENV2 NS3hel (PDB code: 2BMF) (Xu *et al.*, 2005) as search probes. Refinement cycles carried out using REFMAC5 (Collaborative, 1994) were interspersed with model rebuilding using Coot (Emsley and Cowtan, 2004) at the computer graphics. TLS refinement was introduced in the last refinement steps. Three TLS groups are included: NS3 (residues 20-180) & NS2B (residues 49-63), NS3 (residues 181-315), and NS3 (residues 316-618), based on the analysis via TLSMD server (Painter and Merritt, 2006). (B) Conformation II (PDB code: 2WHX). The crystal contains one monomer per asymmetric unit with an estimated solvent content of 51.53 % based on a V_m value of 2.56 (Matthews, 1968). The structure was solved by molecular replacement with the program PHASER (Storoni *et al.*, 2004), using the DENV4 NS3pro (PDB code: 2VBC) (Luo *et al.*, 2008a) and DENV4 NS3hel (PDB code: 2JLQ) (Luo *et al.*, 2008b) as search probes. Refinement cycles carried out using REFMAC5 (Collaborative, 1994) were interspersed with model rebuilding using Coot (Emsley and Cowtan, 2004) at the computer graphics. TLS refinement was introduced in the last refinement steps. Three TLS groups are included: NS2B (residues 49-63), NS3 (residues 20-179), NS3 (residues 180-315), NS3 (residues 317-481), and NS3 (residues 482-618), based on the analysis via TLSMD server (Painter and Merritt, 2006). The quality of the structures were analyzed using PROCHECK (Laskowski *et al.*, 1993). A summary of the refinement statistics and stereochemistry analysis is given in **Table 2. 2**. Solvent accessible surfaces areas were calculated using CCP4 program AREAIMOL with a 1.4 Å radius sphere as a probe. Superposition of structures was carried out using program LSQKAB from the CCP4 suite. Figures were prepared using the program Pymol (DeLano, 2002).

NS3hel (residues 172-618). The native NS3hel structure was solved by molecular replacement with the program Phaser (McCoy, 2007) using the DENV2

NS3 helicase (PDB code: 2BMF) (Xu *et al.*, 2005) as a search probe. The complex with RNA₁₂ was solved with the program Phaser, using refined DENV4 NS3 helicase as the search probe. Refinement cycles were carried out using REFMAC5 (Murshudov *et al.*, 1997) with the TLS (translation, liberation, screw-rotation displacement) refinement option towards the end, using the four following groups for each independent molecule: subdomain 1 (residues 168-326), subdomain 2 (327-481), subdomain 3 (482-618) and the RNA ligand. Refinement cycles were interspersed with model rebuilding using Coot (Emsley and Cowtan, 2004) at the computer graphics. The quality of the structures was analyzed using PROCHECK (Laskowski *et al.*, 1993). A summary of the refinement statistics and stereochemistry analysis is given in **Table 2. 4**. Superpositions of structures were carried out using program LSQKAB from the CCP4 suite. Figures were prepared using the program Pymol (DeLano, 2002).

2.6 X-ray scattering experiments and data analysis of scNS2B₁₈NS3

X-ray scattering data were collected following standard procedures on the X33 beam line of the EMBL Hamburg Deutsches Elektronen Synchrotron (DESY) using a MAR345 image plate (MarResearch, Norderstedt, Germany). The scattering patterns from scNS2B₁₈NS3 at protein concentrations of 2.5 mg/ml and 10.0 mg/ml were measured using a sample - detector distance of 2.4 m, covering the range of momentum transfer $0.1 < s < 4.5 \text{ nm}^{-1}$ ($s = 4\pi \sin(\theta)/\lambda$, where θ is the scattering angle and $\lambda = 0.15 \text{ nm}$ is the X-ray wavelength). The data were normalized to the intensity of the transmitted beam and the scattering of the buffer was subtracted as background. These difference curves were scaled for concentration which allows evaluation of the molecular mass (MM) by the forward scattering $I(0)$ and the radius of gyration R_g using the Guinier (Guinier and Fournet, 1955) approximation. All data processing steps were performed with the program package PRIMUS (Konarev *et al.*, 2001). The distance distribution function $p(r)$ of the particle was computed by the indirect transform package GNOM (Svergun, 1994). For the MM estimation the forward scattering of a BSA solution was taken as reference.

Two low resolution models for the scNS2B₁₈NS3 molecule were built using the program DAMMIN (Svergun, 1992) and GASBOR (Svergun, 1997) as described

in reference (Armbruster *et al.*, 2004). Both approaches start with an ensemble of densely packed dummy atoms (DAMMIN) or dummy residues (GASBOR) inside a search volume defined by a sphere of diameter D_{max} . Eleven independent GASBOR reconstructions were computed and the models were further analyzed using the packages DAMAVER (Svergun *et al.*, 2001) and SUBCOMP resulting in an averaged model for shape representation.

2.8 Fluorescence Correlation Spectroscopy experiments

FCS was performed at room temperature on a ConfoCor 3 (Zeiss, Jena, Germany) using the ATP- and ADP-analogues EDA-ATP ATTO-647N and EDA-ADP ATTO-647N (ATTO-TEC, Siegen, Germany). The 488 nm laser line of a HeNe633 laser was attenuated to 5 mW and focused into the aqueous solution by a water immersion objective ($40\times / 1, 2$ W Korr UL-VIS-IR, Zeiss). FCS was measured in 50 μ l droplets of the diluted fluorescent derivatives of ATP and ADP, which were placed on Nunc 8 well chambered cover glass. Before usage, the cover glasses were cleaned and afterwards treated with filtered and centrifuged 3% gelatin solution in H₂O, in order to prevent unspecific binding (Hunke *et al.*, 2007). The following filter sets were used: MBS: HFT 488/543/633, EF1: LP 655, EF: None, DBS: None. Out-of-focus fluorescence was rejected by a 90 μ m pinhole in the detection pathway, resulting in a confocal detection volume of around 0.25 fl. Fluorescence autocorrelation functions were measured for 30 sec each with 10 repetitions. Solutions of Cy5 in pure water were used as references for the calibration of the confocal microscope. To analyze the autocorrelation functions of fluorescent nucleotides bound, in part, to scNS2B₁₈NS3 or NS3hel (177-618) alternatively, models with the diffusion time and the triplet state were used for fitting. Triplet lifetime and amplitudes were kept at the pre-measured values in the absence of binding proteins. The diffusion times of fluorescent nucleotides and fluorescently labeled nucleotides was measured independently, and were kept fixed during the fitting of the FCS data. Therefore, the determination of the binding constants required only the calculation of the relative amounts of free nucleotides with the short diffusion time and of the bound nucleotides with the diffusion time of the protein.

Table 2. 1 Data collection statistics for scNS2B₁₈NS3.

Parameters	Conformation I	Conformation II (ADP)
Wavelength (Å)	0.9792	0.9792
Cell parameters (Å), P2 ₁	<i>a</i> =52.86 <i>b</i> =88.24 <i>c</i> =76.67 β =94.20°	<i>a</i> =52.42 <i>b</i> =87.72 <i>c</i> =75.78 β =92.9°
Resolution range (Å)	88.3-3.15 (3.30-3.15)	30.0-2.2 (2.3-2.2)
No. of observed reflections ^a	41,223 (6,137)	139,103
No. of unique reflections	12,236 (1,786)	31,361 (2,025)
Completeness	99.8% (99.8%)	90.4% (58.5%)
Multiplicity	3.4 (3.4)	4.4 (3.0)
<i>R</i> _{merge} ^b	0.14 (0.68)	0.08 (0.30)
<i>I</i> /σ(<i>I</i>)	11 (2.1)	16.5 (2.2)

^aThe numbers in parentheses refers to the last (highest) resolution shell.

^b $R_{merge} = \sum_h \sum_i |I_{hi} - \langle I_h \rangle| / \sum_h \sum_i I_{hi}$, where *I*_{hi} is the *i*th observation of the reflection *h*, while $\langle I_h \rangle$ is its mean intensity

Table 2. 2 Refinement statistics for scNS2B₁₈NS3.

Parameter	Conformation I	Conformation II
Resolution range (Å)	20.0 – 3.15	20.0 – 2.2
Rfactor [§] (%)	20.2 (30.8)	20.8 (22.9)
Rfree [#] (%)	27.7 (41.1)	25.9 (23.9)
Average B Factor (Å ²)	61.0	42.9
Water	40.9	38.8
Ligands (ADP / Mn)	-	41.5 / 28.2
Rms deviations from ideality		
Bond lengths (Å)	0.007	0.014
Bond angles (°)	1.090	1.685
Ramachandran plot		
Residues in most favoured regions (%)	83.2	88.9
Residues in allowed regions (%)	16.0	11.2
Residues in disallowed regions (%)	0.8	0.0
Overall G factor*	-0.03	-0.03
PDB code	2VBC	2WHX

[†] The numbers in parentheses refers to the last (highest) resolution shell.

[§] Rfactor = $\sum ||F_{obs}| - |F_{calc}|| / \sum |F_{obs}|$.

[#] Rfree was calculated with 5% of reflections excluded from the whole refinement procedure

* G factor is the overall measure of structure quality from PROCHECK (Laskowski *et al.*, 1993).

Table 2. 3 Data collection Statistics for NS3hel

Parameters	Native	ADP	RNA ₁₂	RNA ₁₂ -AMPPNP	RNA ₁₂ -VO ₄	RNA ₁₂ -ADP-PO ₄	RNA ₁₂ -ADP	RNA ₁₃
Space group	P2 ₁ 2 ₁ 2 ₁	P4 ₃	C2	C2	C2	C2	C2	C2
Cell parameters a,b,c (Å)	a=52.8 b=92.6 c=103.0	a=b=52.3 c=192.9	a=132.0 b=105.5 c=72.5	a=132.8 b=105.2 c=72.6	a=132.5 b=105.2 c=72.4	a=131.7 b=105.6 c=72.7	a=133.0 b=104.8 c=73.0	a=132.4 b=104.6 c=72.3
α,β,γ (°)	α=β=γ=90	α=β=γ=90	α=γ=90 β=117.5	α=γ=90 β=117.9	α=γ=90 β=117.6	α=γ=90 β=117.7	α=γ=90 β=115.9	α=γ=90 β=117.7
Resolution range (Å)	21.2-1.7 (1.8-1.7)	35.4-2.2 (2.3-2.2)	29.8-2.0 (2.1-2.0)	64.3-1.9 (2.0-1.9)	78.3-2.2 (2.3-2.2)	40.8-2.4 (2.5-2.4)	31.9-2.2 (2.3-2.2)	43.4-2.6 (2.7-2.6)
Observed reflections ^a	445,134 (59,415)	176,639 (24,346)	352,858 (44,388)	294,791 (15,191)	127,207 (18,691)	188,211 (24,876)	193,997 (27,739)	119,823 (16,942)
Unique reflections	59,272 (8,383)	24,904 (3,512)	54,798 (7,126)	60,795 (7,262)	44,015 (6,458)	34,344 (4,870)	45,701 (6,666)	26,932 (3,907)
Completeness (%)	99.5 (97.8)	98.3 (95.0)	96.5 (86.4)	87.7 (72.6)	98.8 (99.7)	99.6 (97.5)	100 (100)	100 (100)
Multiplicity	7.5 (7.1)	4.1 (3.8)	6.4 (6.2)	4.8 (2.1)	2.9 (2.9)	5.5 (5.1)	4.2 (4.2)	4.4 (4.3)
R _{merge} ^b (%)	4.5 (75.8)	11.5 (69.9)	8.7 (37.7)	14.4 (41.4)	10.3 (49.7)	14.0 (72.3)	8.3 (58.1)	15.2 (57.9)
Mean I/σ(I)	22.4(2.7)	11.1(1.3)	17.4(4.2)	11.5 (1.9)	11.5(3.3)	13.8(1.7)	13.3 (2.3)	11.0 (1.9)
Monomers in a.u.	1	1	2	2	2	2	2	2
Ligands in a.u.								
ssRNA	0	0	2	2	2	2	2	2
AMPPNP	0	1	0	2	0	0	0	0
ADP	0	0	0	0	2	2	2	0
VO ₄	0	0	0	0	2	0	0	0
PO ₄	0	0	0	0	0	0	0	0
Mn ²⁺	0	1	0	2	2	2	2	0

^aThe numbers in parentheses refers to the last (highest) resolution shell.
^b $R_{merge} = \sum_i ||I_{hi} - \langle I_i \rangle| / \sum_i I_{hi}$, where I_{hi} is the i th observation of the reflection h , while $\langle I_i \rangle$ is its mean intensity.
a. u. asymmetric unit.

Table 2. 4 Refinement statistics for NS3hel

Parameters	Native	AMPPNP	ADP	RNA ₁₂	RNA ₁₂ - AMPPNP	RNA ₁₂ -ADP- VO ₄	RNA ₁₂ -ADP- PO ₄	RNA ₁₂ -ADP	RNA ₁₂ -ADP	RNA ₁₃
Resolution range (Å)	20.1-1.7	19.7-2.0	19.8-2.2	20.0-2.0	20.0-1.9	20.0-2.2	20.0-2.4	20.0-2.2	20.0-2.2	20.0-2.6
No. of reflections (Working/test)	56,198	29,652	23,586	51,971	57,632	41,743	32,553	43,344	43,344	25,521
Rfactor (%)	2,989	1,555	1,275	2,779	3,105	2,213	1,726	2,304	2,304	1,348
Rfree (%)	21.1 (44.1)	20.1 (29.6)	18.6 (21.6)	19.7 (22.1)	19.0 (25.0)	19.6 (30.7)	19.9 (27.8)	19.2 (26.3)	19.2 (26.3)	19.9 (28.3)
No of non-H atoms	23.4 (46.6)	25.1 (38.3)	24.0 (27.1)	24.2 (29.1)	24.8 (33.0)	24.5 (38.3)	25.5 (33.3)	23.8 (35.6)	23.8 (35.6)	27.0 (36.1)
Protein	4,030	3,918	3,854	7,898	8,296	7,977	7,942	7,914	7,914	7,564
Ligands	3,675	3,575	3,569	3,605/3,603	3,622/3,622	3,609/3,611	3,604/3,584	3,609/3,602	3,609/3,602	3,591/3,572
Water	-	32	28	128/128	160/160	185/185	181/181	166/166	166/166	128/128
Mean B-factor (Å ²)	348	311	242	410	718	386	383	357	357	129
Main chain	28.5	33.2	32.3	24.2	25.2	21.6	35.6	35.3	35.3	33.3
Side chain	27.0	32.1	31.7	23.2/23.2	23.6/23.5	20.4/20.4	35.1/35.2	34.8/34.9	34.8/34.9	32.8/32.7
Ligands	28.4	33.1	32.2	24.4/24.4	25.0/24.6	21.4/21.3	35.5/35.4	35.1/35.1	35.1/35.1	33.2/33.1
ssRNA										
AMPPNP		33.5		28.3/28.8	26.1/24.3	29.8/29.4	45.0/46.7	36.2/38.9	36.2/38.9	45.3/45.3
ADP			36.2		23.9/25.1	36.0/38.4	36.9/33.5	34.4/35.9	34.4/35.9	
VO ₄						27.0/26.8				
PO ₄							47.6/47.9			60.7/60.5
Mn ²⁺							24.2/28.1			
Water	37.7	27.6	36.8		14.5/18.1	17.4/17.7	33.7	27.5/26.0	27.5/26.0	
Rms deviations		40.6	37.6	30.6	35.4	27.1		36.9	36.9	32.5
Bond lengths (Å)										
Bond angles (°)	0.002	0.003	0.007	0.011	0.010	0.009	0.007	0.009	0.009	0.008
Ramachandran plot	0.410	0.506	0.981	1.329	1.338	1.223	1.134	1.218	1.218	1.212
Most favoured (%)	91.3	90.0	91.9	92.6	91.4	92.1	91.9	90.7	90.7	90.3
Allowed regions (%)	8.7	10.0	8.1	7.3	8.4	7.9	8.1	9.2	9.2	9.7
Overall G factor	0.13	0.10	0.16	0.07	0.07	0.09	0.12	0.11	0.11	0.10
PDB code	2JLQ	2JLR	2JLS	2JLU	2JLV	2JLX	2JLY	2JLZ	2JLZ	2JLW

Chapter 3 Results and Discussion

3.1 Crystal structures of full length NS3 protein from DENV4

3.1.1 Expression of scNS2B₁₈NS3 and sensitivity to proteolysis

The NS2B cofactor, which is predicted to contain three membrane associated regions, plays an important scaffolding role for the formation of the overall protease structure. In its absence, recombinant proteins are poorly soluble (Keller *et al.*, 2006; Murthy *et al.*, 1999; Sampath *et al.*, 2006; Umareddy *et al.*, 2006; Xu *et al.*, 2006). Apart from its structural role, the C terminal region (residue 67-96) of the internal hydrophilic domain of NS2B cofactor plays a role in catalysis but was not included in the present construct in order to obtain a more stable protein. Residues 49 to 66 were fused to the full length NS3 protein through a flexible Gly₄SerGly₄ linker (**Figure 3. 1A**). This engineered molecule (scNS2B₁₈NS3), which contains the complete NS3 polypeptide chain but only 18 residues out of the 40 from the NS2B cofactor region that are necessary to obtain a fully proteolytically competent enzyme, was used for subsequent structural studies (**Figure 3. 1B**). In spite of a well preserved fold for its protease component, this engineered molecule is devoid of protease activity as shown previously (Erbel *et al.*, 2006). However, after purification, the recombinant scNS2B₁₈NS3 fusion protein remains sensitive to degradation especially at room temperature (**Figure 3. 1C/D**). The three major stable fragments obtained through limited proteolysis were subjected to N terminal amino acid sequencing and mass spectrometry. The closest mapping onto the scNS2B₁₈NS3 protein sequence gives the helicase domain (residue 171~618) as the major product (**Figure 3. 1C/D**). The two minor fragments comprise residues 13~618 and 249~618 of scNS2B₁₈NS3 (**Figure 3. 1C/D**). Appendix B summarizes the results for N terminal amino acid sequencing. Traces of *E. coli* proteases or of thrombin, not eliminated during the purification process, are likely to be responsible for this activity.

Chapter 3 Results and Discussion

(A)

(B)

(C)

(D)

Figure 3. 1 Expression, purification and characterization of scNS2B₁₈NS3. (A) scNS2B₁₈NS3 protein construct. The three membrane-associated regions within the NS2B proteins are represented as filled boxes. A central hydrophilic fragment spanning residues 49-66 of the NS2B protein was linked to the full-length NS3 protein through a Gly₄SerGly₄ linker. Evolutionary conserved residues involved in enzymatic activities in NS3 are indicated. (B) Protein purity checked by SDS-PAGE and by matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry. (C) Protein stability and sensitivity to proteolysis. Freshly purified scNS2B₁₈NS3 protein was incubated at room temperature and examined after different incubation times (hours). DENV2 NS3hel (residues 168-618, ~ 51.5 kDa) is loaded for comparison. (D) MALDI-TOF mass spectrometry examination of scNS2B₁₈NS3 proteolysis. The helicase domain (residue 171~618) is the major product corresponding to the peak (~ 51.3 kDa). The two minor fragments comprise residues 13~618 (~68.3 kDa) and 249~618 of scNS2B₁₈NS3 (~42.2 kDa).

3.1.2 Overall structure of scNS2B₁₈NS3: Conformation I

A summary of the crystallographic data collection and refinement statistics is shown in **Tables 2.1** and **2.2**. A total number of 615 residues are visible in the refined electron density map. Nine residues corresponding to the Gly₄SerGly₄ linker, three residues at the C terminus of the NS2B cofactor and eighteen residues at the N-terminus of the protease domain are missing from the refined model. A ribbon diagram of the scNS2B₁₈NS3 Conformation I is shown in **Figure 3. 2A**. The scNS2B₁₈NS3 molecule adopts a rather elongated shape with approximate overall dimensions of 100 Å × 60 Å × 40 Å. The N terminal protease domain resides next to the entrance of the ATPase active site between helicase sub domains 1 and 2. The main contacts realized between the protease and helicase domains are detailed in **Figure 3. 2A**. They involve the interdomain (linker) region 169 ~179 and two loops that encircle the entrance to the ATP binding pocket. Of note, the rather compact arrangement adopted by the scNS2B₁₈NS3 molecule does not preclude access to the ATP binding site by nucleotides as shown in **Figure 3. 2B**. A basic pocket is formed at the interface between the helicase and protease domains corresponding to the NTPase/RTPase active site. Conversely, if one rotates the molecule by 180° around a

vertical axis, an excess of negative charges becomes visible. Hence, by abutting, the two domains create a charge distribution that is likely to optimize the diffusion of nucleotides substrates towards the ATP hydrolysis site, through long range electrostatic interactions. It is tempting to correlate this structural feature with the 10 fold increase in nucleotide binding affinities observed between scNS2B₁₈NS3 and the isolated helicase domain (see below). Comparison of the scNS2B₁₈NS3 molecule with scNS3-NS4A from HCV (Yao *et al.*, 1999) highlights a major difference for the relative orientations between the helicase and protease domains in the two proteins (**Figure 3. 2C**). The C terminal end of NS3 from HCV participates actively in complex formation with the protease domain, being inserted into the active site of the protease domain as *cis* cleavage occurs at the NS3-NS4A junction. The scNS2B₁₈NS3 molecule from Dengue, on the other hand, has its C terminus located far away from the protease active site (same for the Conformation II, Section 3.1.3.) (**Figure 3. 2C**). As a result, large domain motions would be required to allow cleavage at the NS3-NS4A junction in *cis* by the flavivirus NS3 protein.

(A)

(B)

(C)

Figure 3. 2 Conformation I of the scNS2B₁₈NS3 protein from DENV4. (A) Ribbon representation of the scNS2B₁₈NS3 structure. Secondary structure elements are colored in cyan (α helix) and magenta (β strand). The three sub domains of NS3hel are numbered. NS2B₁₈ which forms a β strand is red. The region linking the protease and helicase (residues 169~179) is in green. Key residues for NS3 enzymatic activities are shown as sticks and labeled. N-terminal residues are also labeled. A

close-up view of the interface between the helicase and protease domains is also shown. (B) View of the electrostatic surface of scNS2B₁₈NS3 (in the right panel, the molecule is rotated by 180° around a vertical axis). Positive potentials are blue and negative red. (C) Side by side comparison of scNS2B₁₈NS3 Conformation I from DENV4 and scNS3-NS4A from HCV (Yao *et al.*, 1999) with their helicase domains oriented similarly. The different domain arrangement is visible between the two structures. The cofactors (NS2B₁₈ and NS4A) are in red and the inter-domain linkers are in green. The N and C terminal residues from the two proteins are labeled.

3.1.2.1 The protease domain

The DENV4 protease domain from the scNS2B₁₈NS3 molecule can be superimposed with the corresponding isolated domain from DENV2 (NS2B₄₇NS3Pro) (Erbel *et al.*, 2006) with an r.m.s. deviation of 0.86 Å for 125 Cα atoms (**Figure 3. 3B**). Thus, the present protease core domain which is inactive retains a conformation similar to a fully active flaviviral protease. At 3.15 Å resolution, its catalytic triad (residues His-51, Asp-75 and Ser-135) appears to adopt the same orientation and residues from the NS2B cofactor (residues 49–66) follow the same path (**Figure 3. 3B**). This confirms the crucial role for the protease activity played by residues 67 to 80 from the NS2B cofactor as Erbel and colleagues inferred from the structure of the WNV protease in complex with a substrate-based inhibitor (Erbel *et al.*, 2006). This segment contributes to the formation of the P2 residue binding site through a large movement induced upon substrate binding (Erbel *et al.*, 2006). Comparison with the NS3 protease from HCV reveals a similar overall conformation including for its NS4A cofactor which contributes one strand to the amino terminal β barrel of NS3 (**Figure 3. 3A**). A significant departure from the isolated protease structure is observed in its C terminal region, with a large reorientation of two β hairpins (**Figure 3. 3A**). In the DENV4 scNS2B₁₈NS3 molecule, this movement projects residues C-terminal from Thr-166 towards the helicase domain. Instead, in DENV2 NS2B₄₇NS3Pro, a hydrophobic core is formed by the clustering of the same β-hairpin B2B-C2 (residues Ile-132 and Phe-116) and E2b-F2 (residue Val-162) with no space left to accommodate the NS2B β2–β3 hairpin which occupies this space during activation of the protease (**Figure 3. 3A**, panel WNV) (Aleshin *et al.*, 2007; Erbel *et al.*, 2006). In our structure, the loop E2b-F2 adopts a slightly more open

Chapter 3 Results and Discussion

conformation possibly mimicking an intermediate state during the protease activation pathway. Several residues including Phe-116, Ile-123, Tyr-161 and Val-162 are evolutionary conserved in flaviviruses. This pocket could thus be targeted for the design of an anti-viral “allosteric” drug that would hinder this dynamic transition.

(A)

(B)

Figure 3. 3 Comparison of flaviviridae protease domains. (A) Ribbon representations for DENV2 NS2B₄₇NS3pro (PDB code: 2FOM), DENV4 scNS2B₁₈NS3 (PDB code 2VBC), WNV NS2B₄₇NS3 (PDB code: 2FP7) and HCV NS3-NS4A (PDB code 1CU1) protease domains. The color scheme for the protease

domains and their cofactor are cyan and red respectively. The bound WNV NS3 protease inhibitor (Bz-Nle-Lys-Arg-Arg-H) is shown as yellow sticks. Two β hairpins (B2B-C2 & E2b-F2) at the C terminal NS3 protease domain and hairpin ($\beta 2$ - $\beta 3$) from the NS2B cofactor are labeled. The side chain of conserved Phe-116, Ile-123, and Val-162 are labeled and presented as purple sticks. The backbone positions of the three catalytic residues (His-51, Asp-75 and Ser-135 in DENV4) are in orange. (B) Superimposition of the C α traces of DENV4 scNS2B₁₈NS3 (blue), DENV2 NS2B₄₇NS3pro (yellow) and WNV NS2B₄₇NS3 (red) (Erbel *et al.*, 2006). NS2B cofactors were removed for clarity. The relative position of the catalytic triad is strictly conserved (green sticks). The main deviation is in the C-terminal region of the protease domain, where two β hairpins (B2B-C2 & E2b-F2) adopt a more open conformation in DENV4 as compared to the closed conformation observed in DENV2.

3.1.2.2 The helicase domain

A comparison of the helicase domain with the isolated DENV2 helicase structure (Xu *et al.*, 2005) returns a value of 0.78 Å for 400 equivalent C α atoms (**Figure 3. 4**, left panel). Structural variations are limited to two segments that form the interface with the protease in the full length NS3: the phosphate binding P-loop at the ATPase active site and a segment comprising residues Arg-460 to Gln-471 in sub domain 2 (**Figure 3. 4**, left panel). Within sub domain 1 of the helicase, a segment of ten residues Gln-243 to Thr-252 which were disordered in the isolated DENV2 helicase structure are now visible. These residues form a β strand that runs antiparallel to a β strand from the helicase sub-domain 2 of a neighboring molecule. Superimposition of the C α trace of scNS2B₁₈NS3 with YFV NS3hel (PDB code: 1YMF) gave r.m.s. deviations of 1.63 Å for 388 C α atoms (**Figure 3. 4**, right panel). Within sub domain 2, residues Glu-396~Lys-405 are disordered in the YFV helicase (Wu *et al.*, 2005), while the equivalent residues in scNS2B₁₈NS3 fold into an α helix. The bound ADP molecule seen in the YFV helicase structure (Wu *et al.*, 2005) can also be accommodated in the ATPase active site of our full length NS3 without steric hindrance, suggesting that the conformation observed in the crystal structure is compatible with ATP hydrolysis. Notice that the ADP is only weakly held by the P-

loop from YFV NS3hel through its β -phosphate. There is no divalent metal ion involved in the interaction. Sub domain 2 and 3 are more close to each other in scNS2B₁₈NS3, suggesting potential intra molecular movements. Such conformational flexibility might be related to the helicase activity.

Figure 3. 4 The helicase domain. Left: α traces for DENV4 scNS2B₁₈NS3 (in blue with the NS2B₁₈ cofactor in red) with DENV2 NS3hel in green (PDB code: 2BMF) (Xu *et al.*, 2005). The P loop and the loop connecting Arg-460 to Gln-471 in sub domain 2 which are involved in contacts with the protease domain, adopt different conformations. Right: Superposition with YFV NS3hel (in yellow, PDB code:1YMF) (Wu *et al.*, 2005). A large difference occurs within sub domain 2: residues between Glu-396 and Lys-405 of the YFV helicase are disordered, while the equivalent residues in scNS2B₁₈NS3 fold into an α helix. The ADP moiety which was weakly bound to the YFV helicase is shown as sticks.

3.1.3 scNS2B₁₈NS3: Conformation II

During the course of improving crystal quality, the discovery of the second conformation of scNS2B₁₈NS3 was unexpected. Conformation II crystals (with similar crystal form to that of Conformation I) (**Figure 2.1A**) were obtained under the same crystallization condition after we changed the protein expression and purification protocol. In the initial attempt of using Conformation I as a searching model in molecular replacement, the helicase domain, but not the protease, was

Chapter 3 Results and Discussion

successfully located. Following the failure to locate the protease domain, the helicase domain and protease domain from the Conformation I were treated as two separate searching probes in the second round of molecular replacement. This molecular replacement strategy successfully resulted in a correct structure solution. Similar to Conformation I, Conformation II adopts an elongated shape and the domain arrangement can still fit into the envelope determined ab-initio using small angle X-ray scattering experiments in solution. It is surprising that the relative orientation between the protease and helicase domain has changed drastically. In contrast to the domain orientation in the previous structure, there is 161° rotation of the protease domain relative to the helicase domain (Figure 3. 5) (Hayward and Lee, 2002). The structures of individual domains are relatively intact, with r.m.s.devations of 1.10 Å for protease domain and 0.97 Å for helicase domain respectively. The inter-domain contacts between the subdomain II of the helicase and the protease domain were completely lost, leaving the ATPase pocket open.

Table 3. 1 Relative orientation of the protease domain.

Property	Value
Fixed Domain	Helicase
Moving Domain	Protease
Rotation Angle	161.0°
Translation	1.8 Å
Closure (Conformation II vs I)	7.6%
Bending Residues	168 - 180

The table is generated based on the output from DynDom (Hayward and Lee, 2002).

(A)

(B)

Figure 3. 5 The relative orientation between the protease and helicase domain of scNS2B₁₈NS3. (A) Side by side view of the two structures of scNS2B₁₈NS3. Secondary structure elements are in cyan (α helix) magenta (β strand). NS2B₁₈ which forms a β strand is in red. The region linking the protease and helicase (residues 169~179) is in green. N terminal residues are also labeled. (B) Superposition of the helicase domain of the two structures. Color codes: NS2B, red; Conformation I, magenta; Conformation II, green. Closest distance between ADP base and Glu-66 of the Conformation I protease domain is labeled.

ADP-Mn²⁺ were successfully soaked into this second crystal form, and yielded a binary complex structure which is almost identical to that found in NS3hel-ADP binary complex (r.m.s. deviations of 0.74 Å for 437 Cα) (Section 3.2.5, Figure 3. 12 right panel). The closest distance between the bound ADP and the protease domain is over 6Å. Interestingly, when the ADP-Mn²⁺ is docked into scNS2B₁₈NS3 Conformation I, the adenine base is very close to the protease domain. The distance between C2 atom of the Adenine ring to Glu-66 is 2 Å (Figure 3. 5B). In the helicase domain, P-loop (Motif I) and a segment comprising residues Arg-460 to Gln-471 (Motif VI) display different conformations in the two structures. Interestingly, the higher-than-average temperature factor value for residues at the P-loop region is greatly lowered down to the average level upon nucleotide binding (Figure 3. 8C). Apparently, Conformation I has to undergo some local structural rearrangements in order to accommodate any nucleotide. The structural flexibility of scNS2B₁₈NS3 might thus be coupled to (or regulated by) its nucleotide binding states.

Table 3.2 summarizes the crystallographic contacts present in the two crystal structures. The helicase domain dominates molecular packaging within the two crystals; in contrast, the protease domain contributes less. It is thus not surprising to observe that the protease domain has the higher-than-average temperature factor in both structures (Figure 3. 8C).

Table 3. 2 Crystallographic contacts in scNS2B₁₈NS3 Conformation I & II.

Number of Contacts between symmetry related molecules, excluding water molecules	Conformation I	Conformation II
scNS2B ₁₈ NS3, 647 residues	46 (0.071) ¹	62 (0.096)
Protease domain, 197 residues	5 (0.025)	12 (0.061)
² Helicase domain, 438 residues	41 (0.094)	50 (0.114)

¹ The numbers in parentheses refers to average number of contacts per residue.

² In total, 22 pairs of contact are exactly in common between neighboring helicase domains in the two structures, none from protease domain. See Appendix D

3.1.4 Nucleotide binding properties of scNS2B₁₈NS3 and of its isolated helicase domain

Using fluorescence correlation spectroscopy (FCS), we measured ATP- and ADP-binding by scNS2B₁₈NS3 and NS3hel respectively. The FCS technique is a

highly sensitive tool to determine binding/dissociation equilibrium constants in the nanomolar range. It monitors the random motion of fluorescently labelled molecules inside a defined volume element irradiated by a focused laser beam. These fluctuations provide information on the rate of diffusion or diffusion time of a particle and this, in turn, is directly dependent on the particle's mass. As a consequence, any increase in the mass of a biomolecule, e.g. as a result of an interaction with a second molecule, is readily detected as an increase in the particle's diffusion time (Bacia and Schwille, 2007; Muller et al., 2003). The characteristic diffusion time, τ_D , was measured by binding labeled nucleotides to non-labeled proteins. The addition of scNS2B₁₈NS3 resulted in a significant change of the mean diffusion time τ_D . This increase is due to the larger mass of the diffusing particle, once ATP ATTO-647N is bound to scNS2B₁₈NS3. This confirmed that nucleotides are bound to scNS2B₁₈NS3 in the presence of Mg²⁺. The percentage of bound nucleotides was analyzed using a two-component binding scheme, $[E]+[S] \leftrightarrow [ES]$. The concentration-dependent binding of scNS2B₁₈NS3 and NS3hel to ATP-ATTO-647N and ADP-ATTO-647N is shown in **Figure 3. 6**. The data in **Table 3. 3** indicate that the affinity of scNS2B₁₈NS3 for both nucleotides is approximately 10 fold higher compared to NS3hel. Interestingly, in the case of the HCV NS3 protein, where the contact between the domains is different, the full length NS3 protein binds ATP more weakly than the truncated helicase domain (Frick *et al.*, 2004; Lam *et al.*, 2004).

Since the helicase activity of the full length DENV NS3 enzyme is about 30 fold higher as compared to the isolated helicase domain (Xu *et al.*, 2005), this supports a major role played by the protease domain to maximize the transfer of the chemical energy derived from ATP hydrolysis to mechanical energy for the strand separation activity of DENV NS3. This also illustrates how viral evolution can operate through the fusion of protein modules (here a chymotrypsin-like protease fold and a helicase domain containing two β - α - β modules) to produce an original molecular solution for a multifunctional enzyme involved in replication.

Figure 3. 6 Fluorescence correlation spectroscopy analysis of the interaction between NS3 and the fluorescent nucleotide-analogues. (A) Concentration-dependent binding of scNS2B₁₈NS3 (■) and NS3hel (□) to ATP-ATTO-647N and (B) Mg-ADP-ATTO-647N linking to scNS2B₁₈NS3 (▲) and NS3hel (Δ). The percentage of bound nucleotides was analyzed using a two component binding scheme, [E]+[S]↔ [ES]. Best fits yielding the binding constants are represented as continuous lines (see text).

Table 3. 3 Nucleotide binding affinity for scNS2B₁₈NS3 and NS3hel.

Protein: Substrate	Dissociation constant (μM)
scNS2B ₁₈ NS3: Mg-ADP ATTO-647N	~ 10-20
scNS2B ₁₈ NS3: Mg-ATP ATTO-647N	~ 20-30
NS3hel : Mg-ADP ATTO-647N	~ 100-200
NS3hel : Mg-ATP ATTO-647N	~ 200-300

The calculation was done with the FCS-Software (AxioVision/Zeiss), Microsoft Excel and a nonlinear curve fit via Origin V.7.5.

3.1.5 Small angle X-ray scattering experiments of scNS2B₁₈NS3 in solution

We used small angle X-ray scattering experiments in solution (SAXS) to assess which quaternary conformations are likely to be present in solution for the scNS2B₁₈NS3 molecule. Solution X-ray scattering (SAXS) patterns from scNS2B₁₈NS3 were recorded and processed as described in “Materials and Methods” to yield the final composite scattering curve in **Figure 3. 7B**. The radius of gyration R_g and the maximum dimension D_{max} of scNS2B₁₈NS3 are 3.9 ± 0.2 nm and 10.1 ± 0.3 nm, respectively, suggestive of a rather elongated particle. Comparison with the scattering curve from reference solutions of BSA yields an estimate for the molecular mass of 74 ± 4 kDa, indicating that scNS2B₁₈NS3 is monomeric at the concentrations

used. A qualitative analysis of the distance distribution function suggests that scNS2B₁₈NS3 consists of a major part yielding a principal maximum in the $p(r)$ at around 3.3 nm (**Figure 3. 7C**) whereas the separated protuberance domain gives rise to a shoulder from 6.1 nm to 10.1 ± 0.2 nm. The gross structure of scNS2B₁₈NS3 was restored *ab initio* from the scattering pattern in Fig. 6A using the shape determination program DAMMIN and the dummy residues modeling program GASBOR as described in “Materials and Methods”. The two approaches yielded similar results but the models provided by DAMMIN could fit the data only up to $s = 0.29 \text{ nm}^{-1}$. In the following, the models obtained with GASBOR are presented, which yield good fits to the experimental data in the entire scattering range (a typical fit as displayed in **Figure 3. 7B**, curve 2, has a discrepancy $\chi = 1.35$). Eleven independent reconstructions yielded reproducible models and the average model is superimposed on the crystallographic Conformation I of scNS2B₁₈NS3 (**Figure 3. 7A**). In solution, the scNS2B₁₈NS3 molecule appears elongated with two distinct domains, a main globular domain having a length of about 6.1 nm, and a hook-like domain of about 4.0 nm in length. The current atomic model for scNS2B₁₈NS3 is well accommodated within the shape of the same protein in solution. The NS3hel and the protease domain lie in the main globular and the hook-like domain, respectively. A very similar elongated SAXS shape for a full length NS3 protein from WNV Kunjin virus was recently proposed (Mastrangelo et al., 2007b). However the fit of individual domains significantly differs from our model. Interestingly, these results suggest that NS3 proteins from different flaviviruses adopt a similar configuration in solution that differs from the NS3 protease-helicase from HCV.

Figure 3. 7 Agreement between the envelope derived from SAXS and the X-ray structure for scNS2B₁₈NS3. (A) Superposition of the low-resolution structure (blue) of scNS2B₁₈NS3 derived from solution X-ray scattering with the crystallographic model (ribbon diagram, yellow). (B) Experimental SAXS curve from scNS2B₁₈NS3 (1) and scattering from typical ab initio model of scNS2B₁₈NS3 (2) Computed by the program GASBOR (Svergun, 1994). (C) The distance distribution function of scNS2B₁₈NS3 was computed from the experimental data by the program GNOM.

3.1.6 On the relative orientation between the helicase and protease domains

A single polypeptide segment spanning residues 169–179 of the NS3 sequence connects the protease and the helicase domains, with most of its residues well defined in the final electron density map (**Figure 3. 8B**). This interdomain (“linker”) region adopts an extended conformation (**Figure 3. 8B**). A sequence alignment of the NS3 protein from the four Dengue serotypes and WNV for the individual protease and helicase domains shows average sequence identities of 66% and 76 % respectively. By contrast, amino acid residues from the linker are less conserved, suggesting limited functional constraints on this region. An exception is the presence in its middle of an acidic residue (Glu-173 in DENV4) followed by a small amino acid (Pro

or Gly) which are found in the four DENV serotypes, in WNV, JEV and YFV (**Figure 3. 8A, Appendix C**)(Xu *et al.*, 2005).

For scNS2B₁₈NS3 Conformation I, attributing the linker region to NS3hel, one obtains a value of 1320 Å² for the total solvent accessible surface area buried between the two domains of the scNS2B₁₈NS3 molecule. This value is comparable to those observed for the formation of stable protein protein interfaces (eg antibody antigen complexes) (Chitarra *et al.*, 1993; Jones and Thornton, 1996). Omitting residues from the interdomain region, the buried surface area between the protease and helicase domain is reduced to only 380 Å². Hydrogen bonds involving residues Ser-62 with Asn-329 and Gly-80 with Arg-202 and Ala-197 are established between the protease and sub domains 1 and 2 of the helicase (**Figure 3. 2A**). For scNS2B₁₈NS3 Conformation II, the helicase and the protease is solely connect by the domain linker. No interface is formed. It was not clear why the same structure was not reproduced. One possible explanation is that, in solution, the two species (probably more) are present in equilibrium and are inter-convertible without much energy cost. During the crystal growth, one species of conformation was favored and packed up into crystal randomly or selected under slightly different buffer conditions. The slight variation in purity of the different batches of the protein may also be a factor that affects the protein crystal growth. The relatively small protease domain contributes very little to the crystal packing, thus it did not affect much the crystal growth and the unit cell parameters (**Table 2.1**). In both structures, average temperature factor for the protease domain is much higher than that of the helicase domain, especially for the domain linker. This suggests that the static or dynamic mobility of the protease domain in the crystal lattice is higher than the helicase, implying there are probably more relative domain orientations than the current resolved two structures (**Figure 3. 8C**).

Mutagenesis studies showed no effect of point mutations on the protease domain on the ATPase activities of the protein (Unpublished data, Xu *et al.*, 2008). It was also shown that an active protease catalytic triad is not required for the stimulatory effect of the protease domain on ATPase activity (Yon *et al.*, 2005). Thus the overall folding of the N-terminal protease domain, the relative domain orientation and movements are probably the factors that contribute to this enhancement.

(A)

DEN1	YVSAIAQAKASQEGPLPE-IEDEVFRKRNLT
DEN2	YVSAIAQTEKSIED-NPE-IEDDIFRKKRLT
DEN3	YVSGIAQTNAEPDGPTPE-LEEEMFKKRNLT
DEN4	YVSAITOAERIGEP-DYE-VDEDIFRKKRLT
WNV	YISAIVQGERMEEP-APAGFEPEMLRKKQIT
	Y Sa I Q e p K T

(B)

(C)

Figure 3. 8 Interdomain linker and relative orientation between the protease and the helicase domains. (A) Sequence alignment of DENV serotypes and WNV scNS2B₁₈NS3 in the linker region (in a rectangle box). Conserved residues are shaded. (B) $|2F_o - F_c|$ electron density map for the linker region between the protease and helicase domains. The map is contoured at 1σ for residues 169~179 which are shown as sticks and labeled. (C) Crystallographic temperature factor analysis of the scNS2B₁₈NS3 structures. For the predicted B factor, blue color codes for the protease domain, green for helicase domain; Black curve displays the observed B factor values. Red circle highlights the P-loop region and yellow circle highlights the domain linker region. Based on the analysis from TLSMD (Painter and Merritt, 2006).

Thus, the inter domain region plays an important but still not clearly defined role for the association of the protease and helicase domains of NS3. Nucleotide binding appears to affect the protein conformation at the interface, and thus the

relative orientation between the two domains (Figure 3. 5B). ssRNA binding triggers global structural rearrangements to the helicase domain (Section 3.2.4, Figure 3. 11B). The effect of RNA binding to the full length NS3 is unknown. Whether the protease domain is directly involved in RNA binding is unknown. A complete analysis of the dynamics of the NS3 protein is desired. Mutagenesis study targeting the domain linker and further structural studies on the full length NS3 structure and its complex with nucleotides and NA will certainly be helpful.

Protease cofactor NS2B is an integral membrane protein that anchors NS2B-NS3 onto the ER derived membrane structures. Section 3.3 provides further discussion about the membrane association of NS2B-NS3 complex and its implications in viral polypeptide processing and genome replication.

3.2 Crystal structures of DENV4 NS3 helicase with its ligands

3.2.1 Structure determination

To decipher the molecular basis underlying the multiple functions of the DENV4 NS3 helicase domain (NS3hel), we determined crystal structures for its apo form to a resolution of 1.7 Å, with a non hydrolysable ATP analog (AMPPNP), with ADP/Vanadate/ Mn^{2+} , with ADP/ P_i / Mn^{2+} , and ADP/ Mn^{2+} , providing snapshots along the complete hydrolytic cycle for a flavivirus ATPase (Table 2. 3 and Table 2. 4). Several structures of NS3hel were resolved in the presence of a bound ssRNA: with either a 12-mer (RNA₁₂) 5'AGACUAACAACU-3' or a 13-mer (RNA₁₃) 5'-UAGACUAACAACU-3', providing a structural basis for sequence-independent RNA recognition and giving insight at atomic resolution into mechanochemical events that couple ATP hydrolysis and RNA unwinding by the flavivirus NS3 helicase.

3.2.2 Comparison of DENV4 NS3hel with other flavivirus NS3 helicases

Superposition with the other flavivirus helicases, including YFV (Wu *et al.*, 2005), WNV (Mastrangelo *et al.*, 2007c), MVEV (Mancini *et al.*, 2007) and DENV2 (Xu *et al.*, 2005), yields r.m.s.devations of 1.77 Å (based on 431 Cα), 1.53 Å (based

on 415 C α , molecule A), 1.05 Å (based on 429 C α) and 0.68 Å (based on 441 C α , molecule B), respectively (**Figure 3. 9**). Among all three domains, domain 1 is the most well-superimposed, the deviant parts reside in domain 2 and 3, especially the two opposite helices, $\alpha 1'$ of domain 2 and $\alpha 7''$ of domain 3. In YFV and MVEV helicases, the distance between this two α helical tips, (C α of Arg-371 and Gln-609 for YFV; C α of Lys-367 and His-605 for MVEV) is 12.5 Å and 7.7 Å (**Figure 3. 9**, inset I), represent the open and close conformations of the ssRNA binding groove, respectively. The distances between corresponding α helices in WNV, DEN2 and DEN4 helicases are 11.8 Å, 10 Å and 8.9 Å, respectively. In addition, when comparing NS3hel of DENV4 with YFV and WNV, another interesting structural variation was a conformational change in domain 2, involving a short helix $\alpha 3'$ and a loop composing residues Thr-409 to Phe-417, which belongs to motif V (**Figure 3. 9**, inset II).

Figure 3. 9 Superposition of Flavivirus NS3 helicases.

Color codes: DENV4 (blue), DENV2 (green), YFV (yellow), MVEV (red) and WNV (purple). Significant differences in sub domain 2 and 3, the opposite helix $\alpha 1'$ (domain 2) and $\alpha 7''$ (domain 3) and motif V are circled. An arrow indicates the open and closure between domain 2 and 3. P-loop shows a great degree of flexibility. Notice DENV4 and MVEV NS3 helicases have the same relaxing conformation while the rest have the closed conformation. Inset I: Comparison of distance between the tip of helix $\alpha 1'$ (domain 2) and $\alpha 7''$ (domain 3) in YFV and MVEV (see text). Inset II:

Conformational changes in motif V. DENV4, DENV2 and MVEV NS3 helicases seem adopt one orientation, while YFV and WNV helicases another. Residues Thr-408 and Arg-418 of DENV4 are labeled.

3.2.3 ssRNA recognition by DENV4 NS3hel

We obtained five isomorphous crystal forms with a 12-mer (RNA₁₂) and one with a 13-mer (RNA₁₃) at a resolution between 1.9 Å to 2.6 Å. A comparison of the contacts established by the ssRNA with NS3hel in its different nucleotide bound states does not reveal any major differences, and RNA recognition appears similar regardless of the ATP bound state, given the resolution limit of this study.

In agreement with earlier observations (Wu *et al.*, 2005; Xu *et al.*, 2005), the ssRNA is accommodated in an extended conformation in the tunnel that separates the RecA-like sub domains 1 and 2 of NS3hel from its sub domain 3 (**Figure 3. 10A/B**). Sub domain 1 binds to the 3' end of ssRNA while the 5' end mainly interacts with sub domain 2 (**Figure 3. 10C**). Both the ssRNA location and polarity follow that of a deoxyuridine octamer bound to the HCV helicase (Kim *et al.*, 1998). The exact path of the sugar-phosphate backbone, however, differs markedly between these two structures, with the ssRNA conforming to one strand of A form dsRNA in the flavivirus helicase, while it is more extended in the HCV helicase. In this respect, the ssRNA binding mode by DENV NS3hel is strikingly similar to that found in eukaryotic DEAD box proteins eIF4AIII or Vasa in complex with polyU, with the RNA moieties closely superimposable between the three structures (see below) (Andersen *et al.*, 2006; Bono *et al.*, 2006; Sengoku *et al.*, 2006). Interestingly, the HCV helicase is more active on DNA templates (Pang *et al.*, 2002), whereas the DENV helicase exhibits higher activity on RNA duplexes (Xu *et al.*, 2005). Five bases of the RNA located at the 5' end are stacked to each other (**Figure 3. 10**). Only the sugar phosphate backbone is well ordered for nucleotides 6 and 7, and only weak electron density is visible for the corresponding bases. Like in the Vasa-ssRNA complex (Sengoku *et al.*, 2006), the RNA structure makes a sharp bend between bases 5 and 6. At the 3' end, the remaining five nucleotides are disordered. No aromatic side chain seems rightly positioned for base stacking as was proposed for HCV or for PcrA (Kim *et al.*, 1998; Soultanas *et al.*, 1999; Velankar *et al.*, 1999). Residues from the

three domains of the protein (but in majority from subdomains 1 and 2) contact the sugar-phosphate backbone either directly or via water molecules (**Figure 3. 10**), explaining how NS3hel can accommodate diverse viral genomic sequences as it translocates towards the 5' end. As seen in the HCV helicase complexed with a dU₈ oligonucleotide (Kim *et al.*, 1998), several evolutionary conserved threonine residues make direct (Thr-244 and Thr-408) or water mediated (Thr-267) contacts to the phosphoryl oxygens of ssRNA (**Figure 3. 10C**). The specificity of DENV NS3hel for RNA should derive in part from several hydrogen bonds that are established with the 2'hydroxyl moieties by carbonyl oxygens of Pro-363 and Pro-233 and by side-chain atoms from Asp-409, and Thr-264 (**Figure 3. 10C**). As expected, RNA recognition appears to largely occur in a sequence independent manner since a 13-mer oligoribonucleotide (RNA₁₃) having a different sequence binds to NS3hel in essentially the same way as the 12-mer (**Figure 3. 10D**). Only the two bases at the 5' end of the 12-mer make hydrophobic and polar interactions with NS3hel. The N2 atom of the guanine at position 2 establishes the sole direct hydrogen bond with the protein via the carboxylic group of Asp-290 (**Figure 3. 10C**). A basic patch (**Figure 3. 10A, inset**) shaped by side chains protruding from Arg-342 and Lys-366 leaves ample space to accommodate the triphosphate moiety of flaviviral genomic capped RNA, an intriguing possibility that might relate to the presence of a strictly conserved AG dinucleotide sequence that is conserved at the 5' end of all flaviviruses genomes.

(A)

(B)

(C)

(D)

Figure 3. 10 The flavivirus NS3h-RNA-AMPPNP ternary complex.

(A) Cartoon representation of NS3hel bound to ssRNA and AMPPNP, a non-hydrolysable ATP analog shown as sticks. The RecA-like sub domain 1 and 2 are in green and cyan respectively. The manganese ion is represented as a yellow sphere. The two insets represent magnified views of the interactions. Top: color coded representation of the NS3hel electrostatic surface potential in the region contacting ssRNA. Residues that contact RNA are displayed as sticks and labeled. Also indicated their motif. Sub domain 3 was omitted for clarity. Bottom: ATP binding site using the same color code for sub domains 1 (green) and 2 (blue), and motif III in yellow. Conserved residues are shown as sticks and labeled, water molecules: red spheres, W-a: attacking water (see text), Hydrogen-bonds: dashed lines. (B) Electrostatic surface view of NS3 ssRNA (RNA₁₂) complex. The large and open ssRNA binding groove is indicated, and the ssRNA is shown as sticks. Notice that the positive charges next to 3' end of the ssRNA on the surface of NS3hel could be the additional RNA binding sites. (C) Schematic representation of the interactions observed between NS3hel and RNA₁₂ in the binary complex. Notice several contacts established with the 2'-hydroxyl groups of the ribose moieties. (D) Interaction network between NS3hel and RNA₁₃: 5'-UAGACUAACAACU-3'. Direct interactions between protein and RNA sugar-phosphate backbone are largely conserved as with RNA₁₂. Residue His-251 in dot-line box belongs to the other NS3hel in the same asymmetric unit. Note Asp-290 remains in contact with Guanine 3 base through a water molecule.

3.2.4 Conformational changes upon ssRNA binding

NS3hel undergoes several major conformational changes upon RNA binding (**Figure 3. 11**). We provide the first description of quaternary changes in the *flaviviridae* RNA helicases that are likely to accompany RNA translocation: Sub domain 3 rotates some 11° away from the ATP binding domains leading to a widening of the RNA binding tunnel and allowing the insertion of several water molecules (**Figure 3. 11B**). Helix $\alpha 2''$ which acts as the hinge for this movement is deformed into a coil. Upon ssRNA binding, we also observe closure between sub domains 1 and 2 by *ca* 12°, leading to a narrower ATP binding cleft (**Figure 3. 11B**). This closure leads to a translation of about 5 Å of the RNA binding motifs within sub domain 2

relative to sub domain 1. This distance, that approximately spans two adjacent bases, might correspond to one translocation step as NS3hel tracks along ssRNA in the tunnel in the 3' to 5' direction (**Figure 3. 11B**). Noteworthy, such pronounced domain rearrangements have not been observed in the HCV NS3 helicase upon oligonucleotide binding (Kim *et al.*, 1998). For DENV4 NS3hel, these movements in the quaternary structure are accompanied by an inward movement of the P-loop leading to a reduction of the volume of the ATP binding cleft and the formation of a salt bridge between Lys-199 and Asp-284 (**Figure 3. 11C**). Thus, in the absence of ATP, allosteric conformational changes triggered by ssRNA binding, convert the P-loop into its substrate-bound conformation, poised to react with ATP without incurring an extra entropic cost.

(B)

(C)

Figure 3. 11 Structural rearrangements upon RNA binding.

(A) R.M.S. deviations between the various NS3hel structures or their individual domains after superposition. (-P-loop) indicates values with residues 195-202 omitted from the calculation. A clear partition can be seen between the RNA-bound (“+RNA”) and the RNA-free NS3hel structures (“-RNA”) suggesting extensive conformational changes upon ssRNA binding which is visualized in Figure 3.11B. No significant

conformational changes are observed for different stages of ATP hydrolysis cycle. The figure is generated by superposing NS3hel or its subdomain structures from different ligand complexes. (B) Comparison between the Apo-NS3hel structure (blue) and NS3hel bound to RNA (red). Top left: Superposition of sub domain 1 highlighting P-loops movements. Left lower panel: Closure of the ATP binding cleft upon RNA binding (both sub domains 1 were superposed to generate this figure). The position of the α -carbon atoms of residues that come in contact with RNA are shown as blue (apo-NS3hel) and red (+RNA) spheres and labeled. Right panel: reorientation of sub domain 3 upon RNA binding (sub domains 2 were superposed). (C) Close up view of the ATP binding site highlighting conformational changes that occur in the P-loop (green) upon RNA binding, in the absence of a bound nucleotide (see text).

3.2.5 The AMPPNP/ Mn^{2+} and ADP/ Mn^{2+} binary complexes

In the apo NS3hel structure, the NTPase active site is filled with solvent molecules and the P-loop adopts a relaxed state with the side chain of Lys-199 pointing outwards (**Figure 3. 11C**). This conformation of the P-loop is also observed for the apo DENV4 scNS2B₁₈NS3 (Conformation I) and is likely to correspond to its resting state in solution, in the absence of either a nucleotide or an RNA ligand. In the AMPPNP/ Mn^{2+} bound state (**Figure 3. 12, left**), the P-loop (residues 192~202) has moved towards the protein core and residues 198~202 become reordered to form an extra turn at the N terminal end of helix α 1. This disorder to order transition is accompanied by an inwards reorientation of Lys-199 and the resulting conformation of the P-loop in the AMPPNP/ Mn^{2+} binary complex closely resembles that shown in **Figure 3. 11C** for the NS3hel-RNA complex or in **Figure 3. 10A** for the ternary complex formed between NS3hel, AMPPNP/ Mn^{2+} and ssRNA (see below). The triphosphate moiety of AMPPNP adopts an extended conformation stabilized by nitrogen amides and the dipolar moment of helix α 1, as well as by contacts with residues from the P-loop, Arg-460 and Arg-463 (Motif VI), Gly 414 and Asn-416, either through direct or water bridged H-bonds. The 3'-OH group of the ribose (C2' endo ring pucker) is hydrogen bonded with the main-chain carbonyl oxygen of Arg-463 and the side chain amide group of Asn-329. The N-7 atom of the adenine base (in an *anti* conformation) forms a weak hydrogen-bond to the ϵ -amino group of Lys-201.

However, the large solvent exposure of the base is consistent with a lack of discrimination between various NTP substrates by NS3 (**Figure 3. 10A**). The ADP/ Mn^{2+} binary complex, shown in **Figure 3. 12 (right)**, reveals a very similar set of interactions as in the AMPPNP/ Mn^{2+} complex, except for a slight movement of Lys-201 leading to a disruption of one hydrogen bond.

Figure 3. 12 The AMPPNP/ Mn^{2+} and ADP/ Mn^{2+} binary complex. Carbon atoms from the adenine base and the ribose were in cyan, and the phosphorus atoms in orange. The manganese ion is represented as a large green sphere and water molecules as red spheres. The candidate of the attacking water molecule for triggering ATP hydrolysis is labeled as W-a.

3.2.6 The ATP hydrolysis cycle

In order to shed light on the ATP hydrolytic cycle, we captured four ternary complexes of RNA-bound NS3h with AMPPNP/ Mn^{2+} , ADP/Vanadate/ Mn^{2+} , ADP/ P_i / Mn^{2+} and ADP/ Mn^{2+} respectively (**Figure 3. 13A**). Several of these complexes could be obtained either by soaking the nucleotide in preformed RNA-bound NS3hel crystals or by co-crystallization of the mixture, yielding similar results. The non-hydrolysable ATP analogue 5'-adenylyl- β , γ -imidodiphosphate (AMPPNP) was used to trap an enzyme substrate ternary complex. In this complex, one water molecule hydrogen-bonded by the side chain atoms of Glu-285 (motif II) and Gln-456

(motif VI) is positioned at right distance for an in-line nucleophilic attack of the γ -phosphate (**Figure 3. 13A, top left**). Interestingly, the geometry of the catalytic water (W-a) relative to the γ -phosphate appears more favorable in the ternary complex (with bound RNA) compared to the binary AMPPNP/ Mn^{2+} complex (**Figure 3. 13B, right panel**): the distance to the γ -phosphate is 3.3 Å and the angle subtended by W-a, the γ -phosphorus and the nitrogen atom of AMPPNP is 170° in the ternary complex whilst the corresponding values are 3.8 Å and 144° respectively for the binary complex. Activation of W-a presumably occurs through proton transfer to Glu-285 and/or polarization by Gln-456 leading to an in-line attack of the γ -phosphate of ATP. As a result, a trigonal bipyramidal pentavalent intermediate is formed. We obtained a model for the structure of this transition state by solving the structure of an ADP/Vanadate/ Mn^{2+} ternary complex (Chen *et al.*, 2006; Smith and Rayment, 1996) (**Figure 3. 13A, top right**). In this structure determined at 2.2 Å resolution, three oxygen atoms and the vanadium atom approximately lie in a plane forming a trigonal planar VO_3 group. The fourth oxygen of vanadate which is in an apical position relative to the VO_3 plane now occupies the position of the attacking water molecule (W-a) and is stabilized by the side-chain of Gln-456. Directly opposite, the other axial position is occupied by an oxygen from the ADP β -phosphate, hence completing the pentacoordination shell of the vanadium atom. The distance between the vanadium atom and the phosphorus atom of the β -phosphate is 2.4 Å (**Figure 3. 13C**). By comparison, the distance between the phosphorus atom of the phosphate group and the β -phosphate is 3.7 Å in the ADP/ P_i / Mn^{2+} ternary complex (**Figure 3. 13A/C**).

The NS3hel protein appears to be active within the crystal environment because a soak with ATP leads to the observation of an ADP- P_i / Mn^{2+} complex which results from ATP hydrolysis (**Figure 3. 13A, bottom left**). Following ATP hydrolysis, both the phosphate and ADP moieties must vacate the nucleotide binding site but we have no direct evidence regarding the exact sequence of these post-hydrolysis events. A small tunnel lined by Pro-195, Ala-316, Thr-317, Pro-326, Ala-455, and Gln-456 could serve as an exit route for the phosphate product of the hydrolytic reaction. In the ADP- P_i / Mn^{2+} complex, the phosphate has moved away from the β -phosphate of ADP and a water molecule is now intercalated between the two species, presumably screening the electrostatic repulsion between their negative charges. In the ADP/ Mn^{2+} ternary complex (**Figure 3. 13A, bottom right**) several

water molecules occupy the location of the phosphate including the position formerly occupied by W-a. A summary of the mechanism proposed for the ATP hydrolysis cycle by the flavivirus helicase is displayed in **Figure 3. 13D**.

(A)

(B)

(C)

(D)

Figure 3. 13 ATP hydrolysis

(A) Snapshots of the ATP binding site during hydrolysis. Upper panel (left to right): the AMPPNP/ Mn^{2+} and ADP/ $\text{VO}_4/\text{Mn}^{2+}$ ternary complexes (with bound RNA). Lower panel: the ADP/ $\text{P}_i/\text{Mn}^{2+}$ and ADP/ Mn^{2+} ternary complexes. The manganese ion and water molecules are pictured as green and red spheres respectively, Hydrogen bonds as dashed lines (see text). (B) Close-up view of the AMPPNP-bound structures. Left: the AMPPNP/ Mn^{2+} ternary complex. An anomalous Fourier map (magenta) was calculated and displayed at a level of 10σ , confirming the presence of the manganese ion displayed as a purple sphere (the next peak are at about 5σ level for the phosphorus, sulfur, and chloride atoms). Also displayed: in green, the difference Fourier map showing water molecules (that were omitted for phase calculation) in the manganese coordination shell (at a level of 3σ), and the $2\text{Fo}-\text{Fc}$ map at a level of 2.0σ in cyan. The attacking water molecule that is activated via interactions with side chain atoms of Glu-285 and Gln-456 is labeled W-a. Hydrogen-bonds are represented by dashed lines and the interatomic distances indicated. Right: Geometrical differences of the catalytic water W-a relative to the gamma-phosphate between the AMPPNP/ Mn^{2+} binary complex ("RNA" with atoms in grey) and the AMPPNP/ Mn^{2+} ternary complex (with bound RNA, in color). (C) Superposition of the four ternary structures from (A), based on the coordination of the nucleotides and Mn^{2+} . (D) The proposed catalytic mechanism for ATP hydrolysis. W-a indicates the attacking water molecule, which gets activated through proton transfer to Glu-285 that acts as a base. The scissile bond is indicated by an arrow and possible exit routes of the hydrolysis product by curved arrows. The hydrogen atoms were modeled based on chemical grounds.

3.2.7 Coordination of the divalent metal ion

Interestingly, during the ATPase catalytic cycle, the coordination of the Mn^{2+} ion appears affected by the presence of ssRNA (**Figure 3. 13B, Figure 3. 14**). In the absence of ssRNA, Mn^{2+} is octahedrally coordinated by oxygen atoms from Thr-200, the γ -phosphate and two water molecules in the equatorial plane and by two oxygen atoms emanating from the β -phosphate and Glu-285 in the apical position.

Conversely, in the presence of ssRNA, Glu-285 no longer directly interacts with the Mn^{2+} ion (the distance increases from 2.3 to 3.5 Å) but instead through an additional water molecule which completes the octahedral coordination shell (**Figure 3. 14**). This change in the coordination shell of the divalent metal ion is likely to affect catalytic activity by shifting the negative charge of the carboxyl group of Glu-285 that acts as the base, making the catalytic water a better nucleophile. Moreover, the alteration of the metal hydration state could also facilitate the release of the ADP/Mn^{2+} leaving group as compared to a leaving group directly coordinated by Glu-285.

Figure 3. 14 Coordination of the divalent ion.

Mn^{2+} ion is shown as green sphere, water molecule red. Glu-285 in gray color is taken from NS3hel-AMPPNP binary complex structure. Distances are labeled.

3.2.8 An evolutionary conserved ssRNA recognition module

Two high-resolution crystallographic structures of DEAD-box eukaryotic proteins involved in mRNA processing, were recently reported: a member of the Exon Junction Complex (EJC) eIF4AIII and the Vasa protein, both complexed with poly-U (Andersen *et al.*, 2006; Bono *et al.*, 2006; Sengoku *et al.*, 2006). Here we report the first structure of a *flaviviridae* helicase in complex with ssRNA. Superposition of these three structures (**Figure 3. 15**) reveals a striking similarity in the molecular details of ssRNA recognition and suggests the existence of an ancestral ssRNA recognition module that was borrowed by the *flaviviridae* during evolution. A

summary of the interactions between ssRNA and conserved residues within the SF2 helicase motifs is presented in **Figure 3. 15**. The two α/β (RecA-like) domains establish extensive interactions with the 2'-OH and phosphate groups of a core of five nucleotides (numbered here from 1 to 5) in a conserved manner (**Figure 3. 15**). Notably, both the conformation and location of the ssRNA with respect to the three helicases DENV NS3hel, eIF4AIII, and Vasa is virtually identical (**Figure 3. 15B**). Beside this conserved core of 5 nucleotides that are well ordered in the protein binding site, both 5' and 3' nucleotides appear disordered, an observation consistent with their susceptibility to RNase digestion for the EJC eIF4AIII (Andersen *et al.*, 2006). Several of these sequence motifs are also present in DNA helicases (Buttner *et al.*, 2007; Jankowsky and Fairman, 2007; Pyle, 2008).

It appears that through fusion with additional domains, this ancestral single strand nucleic acid recognition module has evolved to fulfill different molecular and cellular tasks. In the case of the DENV NS3 protein, both a protease domain and an original protein recognition module have been fused at its N and C-terminal ends respectively.

(A)

	I	Ia	Ic
DENV4NS3	193LHPGAGKTKRIL204	218TLILA ¹ TRVVAAE230	239PIRY ¹ TPAVKS249
eIF4AIII	SQSGTGKTATFS	ALILASTRELAVQ	HACIGGTNVGE
Vasa	AQTGSGKTA AFL	VVIVSTRELAIQ	GIVYGGTSFRH
HCVNS3	APTGSGKSTKVP	VLVLNPSVAATLG	PNIRTGVRTIT
Hel308	MPTAAGKTL LAE	SLYVVPLRALAGE	IGISTGDRDEH
	Ib	II	III
DENV4NS3	259LMCHAF ¹ ETRLLS271	279NLIVMDEAHFTDPC292	311AIFMTATPPGS321
eIF4AIII	VAGTP ¹ RVFDMIR	KMLVLDEADEMLNK	VVLISATLPHE
Vasa	VIATP ¹ RLLDFVD	RFVVLDEADRMLDM	TLMFSA ¹ TFPEE
HCVNS3	TYSTYGKFLADGG	DIIICDECHSTDAT	VVLATATPPGS
Hel308	IVTTSEKADSLIR	SCLVVDEIHL LDSE	VIGLSATAPNV
	IV	IVa/ (QxxR)	V
DENV4NS3	358TVWFV ¹ SIKAGN369	384QLSRKTFDTEYPKT397	406VTT ¹ ISEMGAN416
eIF4AIII	AVIFC ¹ TKRKVD	SMHGDMPQKERESI	ISTDVWARGLD
Vasa	TIVFV ¹ TKRGAD	SIHGDR ¹ LQSQREQA	IATSVASRGLD
HCVNS3	HLIFCHSKKKCD	AYYRG ¹ LDVSVIPTS	VATDALMTGFT
Hel308	VLVFESTRRGAE	FHHAG ¹ LLNGQRRV	VATPTLAAGVN
	VI	Motifs	
DENV4NS3	456QRRGRIGRNPAQ 467	ATP binding/Hydrolysis	
eIF4AIII	HRIGRSGRYGRK	RecA Domain Linker	
Vasa	HRIGRTGRVGNN	2'OH (RNA Specific)	
HCVNS3	QRRGRTGRGKPG	Backbone Pi	
Hel308	QMAGRAGRP GMD		

(B)

Figure 3. 15 An evolutionary-conserved ssRNA recognition module of the DE-A/x-D/H helicases. (A) Structure-based sequence alignment of SF2 helicases discussed in the text. Only the regions encompassing the conserved motifs involved in ATP binding and hydrolysis, interdomain communication, binding of the ribose 2'-hydroxyl of RNA or backbone phosphate is included. Residue numbers are for the DENV NS3. Motifs in italics delineate residues which are structurally equivalent (based on the superposition in panel B) but are not conserved. (B) ssRNA recognition by DEAD/DExH Helicases. Left panel: comparison of ssRNA complexes of eIF4AIII (in green; PDB code: 2J0S and 2HYI), Vasa (in blue; PDB code: 2DB3) and DENV NS3hel (in red, this work). The structures are overlaid based on the ssRNA sugar-phosphate backbone. Protein α -carbon atoms are displayed as smooth ribbon and ssRNA as sticks. $C\alpha$ atoms of the structurally conserved RNA binding residues (belonging to Motif I to V) are marked by colored spheres. Right panel: schematic illustration of the ssRNA recognition modules of the DE-A/x-D/H helicases. For domain 1, motif Ia interacts with 2'OH of Base 3 and Phosphate group of Base 4; motif Ib with 2'OH of Base 4 and Phosphate group of Base 5; motif Ic, less conserved, with Phosphate of Base 5. For domain 2, motif IV interacts with 2'OH of Base 1 and Phosphate of Base 2; motif V with 2'OH of Base 2 (observed in DENV NS3) and Phosphate of Base 3; motif VI, less conserved, with Phosphate of Base 3 and Base 4.

3.2.9 Conformational changes during ATP hydrolysis and RNA unwinding

Based on our crystallographic results, RNA binding appears as the major determinant that dictates the conformation of the flavivirus NS3 helicase (**Figure 3.11**). The NS3h helicase was crystallized in three different crystal forms in the absence of RNA (**Table 2.3**), all of which are closely superposable and RNA binding appears sufficient to convert the helicase to its closed conformation. However, no change in the quaternary structure of NS3hel was observed in the various altered nucleotide states, either for binary or for ternary (RNA-bound) complexes. This was slightly surprising in the light of the unwinding mechanism proposed for the HCV NS3 protein, a DEXH helicase, where the protein cycles between open and closed states depending on the nucleotide-binding state and where the ATP hydrolysis cycle is coupled to loose versus tight RNA binding and translocation (Pyle, 2008). This observation might simply relate to constraints imposed on the NS3hel protein by forces from the crystal lattice (eg in its RNA-bound C2 crystal form) that would restrict conformational changes during the ATP hydrolytic cycle. The fact that NS3hel retains ATPase activity in the C2 crystalline form (ATP can be soaked into the preformed crystal and the ATPase reaction occurs), suggests that both the NS3h-ATP, the NS3hel-Vanadate/ADP/ Mn^{2+} transition-like state and ADP/Pi/ Mn^{2+} structures represent biologically-relevant snapshots. Conversely, the crystal structures we obtained for the ADP/ Mn^{2+} binary and ternary complexes might not be as relevant because at this stage of the reaction, the enzyme is expected both to release the reaction product and translocate along RNA, a step presumably accompanied by several conformational changes (Myong *et al.*, 2007). Thus, at least another conformational state of NS3hel probably exists that could not be trapped in our experiments and which might correspond to one of the conformations predicted for the flaviviridae helicase using normal mode analysis (Zheng *et al.*, 2007).

Two recent structural studies on DEAD box helicases (eIF4AIII and Mss116p) reveal similar observations that different nucleotide transition state analogs are able to bind to the protein without inducing conformational changes (Del Campo & Lambowitz, 2009; Nielsen *et al.*, 2009). One possibility is that the crystal packing forces locked the enzyme in one conformation. Under this circumstance, it is noted that the enzyme is able to hydrolyze ATP and do no work (no conformational changes). It suggests that it might be meaningless to try to work out a number about

how many base pairs unwound per ATP hydrolysis cycle. Rather, studies aiming to study the energy utilization efficiency might be more meaningful. It also suggests that it might be very difficult to catch the enzyme in another conformation (or other conformations, if existing) which distinguishes different nucleotides. Other conformations may well exist and may be very important for the enzyme functioning in solution. On the other hand, they could be very transient and difficult to capture.

3.2.10 Insights into the RNA unwinding mechanism by NS3.

Based on the experimental complexes between DENV NS3hel and ssRNA described above, we performed a simple modeling exercise which consisted in extending the bound ssRNA strand into the 5' direction and pairing it with a complementary strand to form a dsRNA substrate for NS3h (**Figure 3. 16**). Modeling was aided by a crystallographic Hel308 protein-dsDNA structure, a SF2 helicase with a related architecture for both the single-strand nucleic acid binding tunnel and the ATPase active site (Buttner *et al.*, 2007; Wu *et al.*, 2005; Xu *et al.*, 2005). As seen in **Figure 3. 16**, one β -hairpin element of DENV NS3 (joining strands $\beta 4A'$ and $\beta 4B'$, see (Xu *et al.*, 2005)) is likely to play a crucial role for separating the two strands by disrupting base stacking and stabilizing the unwound duplex (Buttner *et al.*, 2007). In this model, one RNA strand would migrate through the tracking ssRNA tunnel, while the other strand is forced towards the back of the protein by the β -hairpin. We note a relative variability of the residues exposed at the surface of the β -hairpin scaffold among several SF2 helicases: in the Hel308 DNA helicase, it consists of a stretch of charged or aromatic residues, of charged residues in HCV NS3 whilst the β -hairpin from Dengue NS3, appears more hydrophobic (**Figure 3. 16**). Interestingly, a topologically equivalent hairpin (christened "separation pin") was also proposed to play an active role in DNA unwinding by UvrD, a bacterial SF1 helicase (Lee and Yang, 2006). Thus this important functional feature is not limited to SF2 helicases but appears to be shared between helicases of these two superfamilies. How residues from the β -hairpin precisely interact with the fork and how this affect RNA unwinding and processivity remains to be determined for a DExH helicase. Moreover, it is not known with certainty whether sub domain 3 of the DExH viral helicase establishes any interaction with dsRNA during the unwinding reaction, an attractive hypothesis

considering that the topologically equivalent subdomains 1B and 2B from SF1 helicase UvrD do interact with the DNA duplex (Lee and Yang, 2006). Recent single molecule studies would favor a second point of contact of the flaviviridae helicase ahead of the fork, presumably by contacting dsRNA (Cheng *et al.*, 2007a). The driving force for strand separation itself remains elusive (Betterton and Julicher, 2005): in an active mechanism, the β -hairpin could function as a “helix opener” by actively disrupting base pairs hydrogen-bonds. Quaternary movements in the three NS3hel sub domains gradually accumulating tension on the protein-NA complex might also assist this process by exerting a force or torque on the duplex, as was suggested by single molecule studies using the HCV helicase (Myong *et al.*, 2007; Pyle, 2008). In a passive mechanism, the β -hairpin would merely act as an “insulator” by taking advantage of the spontaneous opening of the duplex at the fork, through thermal motion, and by stacking with the melted bases. In our model, base +1 would interact with the conserved Leu-443 that projects from the β -hairpin disrupting base stacking and thus splaying apart dsRNA at the fork region (**Figure 3. 16**). We also noticed departure from A-form RNA between bases 5 and 6 at the 3' end of the RNA₁₂ oligomer bound to DENV NS3. A distorted RNA conformation was indeed observed in the Vasa-polyU complex and was proposed to be important for the mechanism of RNA unwinding (Sengoku *et al.*, 2006). Another possibility is that this protein induced strain, however, is not directly related to RNA duplex unwinding but rather to prevent NS3 from sliding backwards along the ssRNA substrate.

Figure 3. 16 RNA unwinding by the NS3 Helicase. Left panel: Proposed model for

Chapter 3 Results and Discussion

a NS3 helicase complexed with dsRNA and ATP. Right panel: close-up view of the proposed helix opener, β -hairpin from domain 2. Conserved hydrophobic residues are shown as sticks and labeled. Partially melted dsRNA is modeled based on comparison between NS3-ssRNA and the archaeal Hel308-dsDNA (PDB code: 2P6R) structure. Color code: subdomain 1, green; subdomain 2, cyan; subdomain 3, orange; β -hairpin, yellow; dsRNA strands are in red (tracking strand) & purple. At the dsRNA fork, three hydrophobic patches protrude from the β -hairpin: patch 1, formed by Pro-431 and Leu-443 stacks with base +1 of the strand bound to NS3; patch 2, Leu-429 and Ala-444, stacks with base 0 and possibly disrupt the base stacking between bases 0 and base +1 of the same strand. Likewise, patch 3, Pro-446 and Ile-447 located opposite to patch 2 and patch 4, Val-432, Leu-434 and Ile-442 to patch 1 stack with the bases 0 and +1 of the other free strand.

The thermodynamics explanation of the energy transfer from ATP hydrolysis to unpairing of the dsRNA is indeed unclear at the moment. It is commonly thought that the chemical energy released from ATP hydrolysis is transferred to mechanical energy which in turn drives the enzyme moving along NA and unwinding NA. This serves as the basis for the mechanism of all helicases. Models of helicase translocating along single stranded and/or unwinding duplex NA are proposed based on studies on different helicases using different approaches which are briefly summarized in **Section 1.3.3.2**. Apparently, there is no single model could explain all phenomena, regardless of variations between different measurement techniques and analysis methodologies. Furthermore, the different helicases might simply evolve to have different molecular mechanisms to perform different molecular tasks in different environments.

For DENV NS3 helicase, the closest helicase that has been extensively studied is HCV NS3. Yet there is still no single model that could explain perfectly how HCV NS3 works. Both positive and passive unwinding models are proposed and each is supported by some experiments (Section 1.3.3.2 and the references). Single molecule kinetic studies are very powerful in studying the strand unwinding speed and step sizes (Dumont et al, 2006; Myong et al, 2007). The fact that DENV NS3 helicase is not a processive helicase and is only an RNA helicase highlights subtle differences

with the HCV NS3 helicase which is more processive and is a DNA/RNA dual functional helicase. The model based on the DENV NS3 helicase ssRNA complex and another DNA helicase Hel308 is proposed to highlight the unknown and the possible ways of unwinding that NS3 could have in a conserved manner. Any further and more detailed models require more data input, for example, a new NS3-dsRNA complex structure or some single molecule kinetic studies.

3.3 Insights into the viral life cycle: Roles that NS3 plays

3.3.1 Membrane association via NS2B

We modeled the scNS2B₁₈NS3 structure onto a planar membrane platform (Fig. 7). Based on secondary structure predictions (Cserzo *et al.*, 1997; Hirokawa *et al.*, 1998) and deletion analysis of NS2B (Falgout *et al.*, 1993), we assume that residues N-terminal from 49 and C-terminal to 96 of NS2B are anchored into the membrane (**Figure 3. 17A**). The NS2B₄₇NS3pro structure was first modeled onto the planar membrane platform, and scNS2B₁₈NS3 was then aligned using the orientation for the protease domain as a guide, yielding a molecule orientated roughly parallel to the membrane (**Figure 3. 17C**). Interestingly, our model places the exposed hydrophobic loop (Gly-29~Leu-30~Phe-31~Gly-32) from the NS3 protease domain towards the membrane, supporting a tripod-like positioning of the NS2B-NS3Pro onto the membrane (**Figure 3. 17B**). This loop which is conserved among flaviviruses was proposed to play a role analogous to the amphipathic N-terminal helix α_0 of the HCV protease (Aleshin *et al.*, 2007; Brass *et al.*, 2008; Yan *et al.*, 1998). N-terminal helix α_0 of the HCV protease was suggested to be the membrane association determinant of NS3-NS4A protease (Brass *et al.*, 2008; Yan *et al.*, 1998). The fact that the NS2B₄₇ could form a “belt-like” structure that wraps around the protease domain to form the active site was intriguing (Erbel *et al.*, 2006). The observation of a similar conformation for a flaviviral NS2B₄₇NS3pro complex containing a trypsin inhibitor aprotinin/BPTI at the active site (Aleshin *et al.*, 2007) has confirmed this unique mode of interaction between a protease and its cofactor. Available structures clearly indicate that the central hydrophilic region of the NS2B cofactor can adopt at least two distinct orientations representing the substrate-bound state and the substrate-free state, respectively. Residues 65~67 on NS2B constitute the hinge around which the C-

terminal region (residues 68~95) swings around to form the active site upon substrate/inhibitor binding (**Figure 3. 17B**). Alternatively the same residues fold back to leave the site accessible for the substrate.

Figure 3. 17 Model for the membrane bound NS2B-NS3 protein complex. The model is generated based on comparisons with the apo-protein structure of DENV2 NS2B₄₇NS3pro (PDB code: 2FOM) and with substrate-bound structure of WNV NS2B₄₇NS3pro (PDB code: 2FP7) (Erbel *et al.*, 2006). (A) Predicted membrane spanning topology of NS2B protein. Essential cofactor (residues 49~96) is tethered to the membrane via the N, C terminal membrane associated region. (B) Membrane association of NS2B-NS3Pro. Two distinct orientations of the hydrophilic region of NS2B cofactor (NS2B₄₇) are possible, representing either the substrate-bound state (red) or the apo state (blue). The hinge region (pointed by a curved block arrow) is centered at residue 66, where flipping of the C-terminal region of NS2B₄₇ towards or away from the protease active site upon substrate binding, occurs. NS3 is represented as a green ribbon. Residues from the catalytic triad His-51, Asp-75, and Ser-135 are shown as sticks and van der waals spheres and labeled. The model places an exposed hydrophobic loop (Gly-29~Leu-30~Phe-31~Gly-32), labeled as “GLFG”, next to the

membrane. Putative membrane association points are indicated as black triangles. Missing residues in the crystal structures are displayed as black lines. (C) Membrane association of NS2B-NS3 (I). The model is generated by superposition of the protease domain of NS2B₁₈NS3 Conformation I to the membrane associated NS2B-NS3Pro. Notice the ssRNA binding groove within helicase domain is facing towards the membrane in the model. (D) Membrane association of NS2B-NS3 (II). The model is generated by superposition of the protease domain of NS2B₁₈NS3 Conformation II (green) to the membrane associated NS2B-NS3Pro. NS2B₁₈NS3 Conformation I is in gray color. (E) Another view of the membrane association of NS2B-NS3 (II). In this case, the helicase domain projects away from the membrane, making an interaction with ssRNA as well as NS5 easier. The model for the membrane (Phosphatidylcholine) was generated using VMD (Humphrey *et al.*, 1996).

The projection of the protease domain is likely to be limited by the tripod-like membrane association (**Figure 3. 17B**). However, the helicase domain could be less restricted via the sole linkage with the protease domain (**Figure 3. 17D/E**). Similar to NS3-NS4A from HCV (Brass *et al.*, 2008; Yao *et al.*, 1999), the membrane association of flavivirus NS2B-NS3 complex and the spatial arrangement of its functional domains could have implications in viral life cycle (Section 3.3.2).

3.3.2 Implications for polyprotein processing and formation of the replication complex

Successful flaviviral infection requires immediate expression of viral proteins and efficient viral replication (Clyde *et al.*, 2006; Westaway *et al.*, 2003). Once into the cell, immediate viral polyprotein translation yields first generation viral polyprotein. During or after initial polyprotein processing, assembly of the virus replication complex starts and first round of viral replication takes place. Moving on, there are more rounds of viral protein translation and genome replication, followed by viral particle packaging, virus maturation and release of infectious viral particles (Lindenbach *et al.*, 2007). Clearly, initial polyprotein processing and the assembly of the viral replication complex for immediate early viral genome replication are critical for the success of a viral infection event (**Figure 3. 18**).

Figure 3. 18 Early events during viral infection. Co/post translational processing of viral polyprotein yields individual functional protein or protein complexes: (1) generation of functional NS2B-NS3 protease itself is probably the first event during viral polyprotein processing; (2) NS3 and NS5 are the two key viral enzymes involved in the formation of viral RC (replication complex); (3) small membrane proteins NS4A/NS4B are suggested to be part of the viral RC, probably by inducing membrane structures and by interacting with viral enzymes and RNA; (4) NS1 is also required for viral replication, although the exact timing of its processing by proteases inside ER lumen is relatively uncertain. The assembly of the viral RC by viral genomic RNA, viral and host protein, is believed to take place on small hydrophobic viral protein induced (ER or ER derived) membrane structures.

3.3.2.1 NS2B-NS3 protease and viral polyprotein processing

During viral infection, the single chain polyprotein is co- and post-translationally processed by viral and host proteases into 10 viral proteins (Lindenbach *et al.*, 2007). This process is vital to provide functional viral proteins or protein complexes for virus genome replication, virus particle assembly, and probably suppression of host immune response (Lindenbach *et al.*, 2007). The NS2B-NS3 protease is responsible for all cytoplasmic cleavage events including protein junctions (NS2A/NS2B, NS2B/NS3, NS3/NS4A, and NS4B/NS5) and protein internal cleavages (C, NS2A and NS4A proteins) (Figure 1. 5).

Several groups have reported the *cis* cleavage of the NS2A/NS2B (Chambers *et al.*, 1990c; Preugschat *et al.*, 1990), NS2B/NS3 (Bera *et al.*, 2007; Chambers *et al.*, 1990c; Preugschat *et al.*, 1990; Wu *et al.*, 2003), and within NS3 (Arias *et al.*, 1993; Bera *et al.*, 2007; Teo and Wright, 1997) by NS2B-NS3 protease from different flaviviruses. The efficiency ranking seems to be NS2A/NS2B > NS2B/NS3 >> NS3458/459. On the other hand, the NS3/NS4A junction was shown to be efficiently cleaved in *trans* and the cleavage is slower than that at NS2A/NS2B/NS3 junctions (Lobigs, 1992; Zhang *et al.*, 1992; Zhang and Padmanabhan, 1993). Based on the structural studies on NS2B₄₇NS3Pro and scNS2B₁₈NS3, we can provide a possible explanation for such a sequential *cis/trans* cleavage pattern around and within NS2B-NS3 protein complex. NS2A/NS2B and NS2B/NS3 junctions are very close to the protease active site (Figure 3. 17). Single chain peptide linkage helps the protease capture the nearby substrate, probably even before it is well folded and is mature. On the other hand, the NS3/NS4A junction is more accessible and thus cleaved by another molecule of NS2B-NS3 protease (Figure 3. 17). Although the possibility of *cis* cleavage at the NS3/NS4A junction cannot be completely ruled out, considerable amount of global structural rearrangement at the domain linker region and the helicase domain is required to bring the NS3/NS4A junction to the protease active site of the same molecule. It is also unlikely that the NS3/NS4A junction cleavage takes place co-translationally while the helicase is not well folded, as *in vivo* study identified uncleaved NS3-NS4A polypeptide (Lobigs, 1992).

For several flaviviruses, a specific *cis* cleavage site of NS3 by the viral protease has been observed corresponding to Arg-458~Gly-459 in helicase sub domain 2 (Bera *et al.*, 2007; Teo and Wright, 1997). Since Arg-458 is buried inside the protein and inaccessible to proteases, cleavage at this position would require large conformational changes, reorientations of the protein domains, and even partial unfolding of the helicase domain. These might be the rate limiting steps that are consistent with the fact that this internal cleavage is less efficient compared to the cleavage at the NS2B-NS3 junction. The biological implications of internal cleavage of NS3 in the flavivirus life cycle are not known, although this site is also cleaved in virus-infected cells (Arias *et al.*, 1993).

NS4B/NS5 junction is readily cleaved by functional NS2B-NS3 protease, which gives rise to individual NS5 protein (Cahour *et al.*, 1992; Chambers *et al.*, 1991;

Falgout *et al.*, 1991; Lin *et al.*, 1993b; Zhang *et al.*, 1992). Internal cleavage near the C terminus of NS4A by the NS2B-NS3 protease is also a prerequisite for the efficient cleavage at the downstream NS4A/NS4B junction (Lin *et al.*, 1993a). Since NS4A and NS4B are implicated in viral replication, probably taking part in the formation of replication complex (See Section 1.2.4.5/6 and the references therein), processing NS4A and NS4B are supposed to be completed before the viral replication (**Figure 3.18**).

It was also shown that the conformation of the non-structural polyprotein precursor NS1 to NS5 affects NS3-NS4A cleavage efficiency in DENV2 (Zhang and Padmanabhan, 1993). Thus an additional level of regulation for viral polyprotein processing could derive from protein-protein interactions and membrane anchorage via the NS2B protein.

3.3.2.2 As part of the viral genome replication complex

3.3.2.2.1 RTPase activity of NS3 and viral genome capping

DENV NS3 has been demonstrated to have RTPase activity, which was suggested to be the first step of virus genome capping (**Figure 3.19A**)(Benarroch *et al.*, 2004; Sampath *et al.*, 2006; Wang *et al.*, 2009; Yon *et al.*, 2005). The RTPase is likely to share the same active site and thus the same chemical reaction with NTPase active site (Benarroch *et al.*, 2004; Wang *et al.*, 2009).

It is shown that ssRNA or dsRNA with 5' overhang of longer than 7 bases is better as RTPase substrate (Wang *et al.*, 2009). Moreover, solvent ionic strength dependant inhibition of ssRNA binding, ssRNA stimulated ATPase activity (but not basal ATPase activity), and RTPase activity suggests that ssRNA binding to the helicase is important in modulating the reactivity of the NTPase/RTPase catalysis. This is in accordance with the structural explanation about the RNA stimulated ATPase activity (Section 3.2).

The 3'OH group of the ATP forms a hydrogen bond with the backbone carbonyl group of Arg-463." The positive patch near the ATPase active site extends to the bottom of the helicase subdomain I (**Figure 3.19B**) which may provide binding surface for the incoming RNA backbone of the 5' end of the viral genome. Due to the limited space at the ATP binding pocket, the stacking interaction between the A1 and

G2 bases of the RNA is to be disrupted in order for the triphosphate group enters the reaction centre.

(A)

(B)

Figure 3. 19 Flavivirus genome 5'cap formation and NS3 RTPase modeling. (A) Four enzymatic modifications are required for cap formation. The 5'RNA triphosphatase has been mapped to the viral NS3, while the N-7 and 2-O MTase activities are performed by NS5; the guanylyltransferase remains to be identified. Phosphates from different molecules are colored individually to indicate their sources. Adapted from Dong *et al.*, 2008. (B) Model of NS3 RTPase. The model is based on NS3hel-RNA₁₂-AMPPNP complex structure. The helicase domain is displayed as electrostatic surface and the protease (Conformation II) is shown as pink cartoon. 3'

hydroxyl group of ANPPNP is labelled and the 5' end of the viral genomic RNA is modelled such that there is no steric hindrance between protein and RNA. Therefore the stacking interaction between the A1 and G2 bases of the RNA could not be maintained (black arrow). The positively charged helicase domain I surface might provide a binding site for the RNA backbone.

Whether the same strand of RNA is required to bind to the same enzyme while it is serving as RTPase substrate? By carrying out RTPase assay in the presence of polyU in excess, the question can be answered. The NTPase/RTPase active site is relatively open and easily accessible, in contrast to that of the DEAD box or SF1 RNA helicases (containing Q motif). It is thus difficult to envisage any specificity towards RNA sequence at the active site. Interestingly, HCV NS3 helicase, sharing a great similarity with dengue at the NTPase active site, lacks significant RTPase activity (Wang *et al.*, 2009). Therefore, there seems to be structural features for ssRNA with 5'-triphosphate to differentiate the active sites between the two enzymes. There might be additional RNA binding site(s) on the surface of the dengue NS3 helicase domain (Li *et al.*, 1999; Sampath *et al.*, 2006). Surface charge variation might promote specific enzyme-substrate pairing. Fine biochemical and structural analysis on these helicases and their interaction with RNA substrate are needed in order to elucidate this specificity. The evolutionary differences in the replication strategies between HCV and dengue may also be partially accountable. HCV viral genome is not capped, thus enzymatic properties related to RNA capping (e.g. RTPase activity from NS3 and MTase activity from NS5 of dengue and other flaviviruses) is absent in HCV.

Capping of the progeny viral RNA is essential for generating infectious viral particles (Dong *et al.*, 2008; Lindenbach *et al.*, 2007). Whether the multistep reaction of capping is completed in concert or separately is still an open question (**Figure 3.19**). The dynamic property of the NS3FL protein and the secondary structures at the viral 5'UTR region further complicated the analysis. The possibility of the interaction between the protease domain and the incoming viral 5' end genomic RNA could not be excluded. Given the limited information about the RTPase activity of NS3 and the virus replication process, a model as in **Figure 3.19B** shall be considered with caution.

3.3.2.2.2 Assembly of the viral replication complex

The assembly of the viral RC is a complex process. It requires correct organization of viral genomic RNA, viral replicases and accessory proteins on cellular (ER or ER derived) membrane structures (Murray *et al.*, 2008). Being different from later replication cycles where number of copies of enzymes and templates are in great excess, (-) RNA synthesis is limited to as few as one copy of essential components for viral RC. Thus specific recognition of viral genomic RNA by viral replicases, namely NS3 and NS5, are essential.

A simple model for DENV (-) RNA synthesis was proposed (Filomatori *et al.*, 2006). The identified promoter at 5'UTR binds the viral RdRp and facilitates template recognition at the 3' end of the genome via genome circularization (Filomatori *et al.*, 2006; Villordo and Gamarnik, 2008). NS3 helicase was proposed to prepare viral RNA to act as template for NS5 RdRp. NS3 thus needs to correctly recognize the viral genome via primary sequence or secondary structures and initiate viral genome reorganization (**Figure 3. 20**). In order to do so, NS3 may work intimately with NS5 and thus be brought over to the sites wherever it is needed. Indeed, NS5 is known to interact with NS3 and regulate its helicase/NTPase activity (Cui *et al.*, 1998; Johansson *et al.*, 2001; Kapoor *et al.*, 1995; Yon *et al.*, 2005). Alternatively, NS3 may also be directly recruited to vRNA. Specific interaction of NS3 with viral RNA has been reported for HCV (Banerjee and Dasgupta, 2001) and JEV (Chen *et al.*, 1997a). In our study, ssRNA with 5'AG di-nucleotides seems to be required for complex crystallization. Interestingly, both the (+) vRNA and (-) vRNA contain strictly conserved 5' AG di-nucleotides. Could NS3 recognize specifically the 5' end of vRNA and then be transferred to the 3' end where it starts translocating at 3'→5' direction? This is very similar to the model for NS5 RdRp recruitment by the vRNA (**Figure 3. 20**). This speculation deviates from the traditional view about RNA helicase, in which helicases are not supposed to possess specificity to primary sequence of RNA. Further experiments are required to examine such a possibility.

Figure 3. 20 Model for DENV minus strand RNA synthesis. The viral genome circularizes in the absence of proteins mediated by 5'-3' UAR and 5'-3' CS hybridization. The viral RdRp binds to a 5' stem loop (SLA), and by long-range RNA-RNA interactions the polymerase is transferred to the site of initiation at the 3' end of the genome. Adapted from Alvarez *et al.*, 2008.

The flexibility of membrane-associated NS2B-NS3 complex (**Figure 3. 17**) suggests a regulation mechanism of the replication complex assembly. Conformational changes that move the helicase away from membrane (**Figure 3. 17D**) could liberate NS3 helicase and increase its accessibility to RNA and NS5. A similar mechanism has been proposed for HCV NS5B RdRp, which by itself is membrane-associated through its C terminal trans membrane domain (Appel *et al.*, 2006). It is also evident that viral RNA replication takes place at ER derived membrane structures (Smooth membrane vesicle like structures, SMS; Also called vesicle packets, VP), different from the likely cellular location for viral polyprotein processing (Convolute membranes (CMs) and paracrystalline structures (PCs)) (Mackenzie and Westaway, 2001; Westaway *et al.*, 2003). In addition, host cell lipid metabolism and intracellular vesicle trafficking related proteins (VAP-A, VAP-B) have been linked to regulation of HCV replication (Appel *et al.*, 2006). In this aspect, flaviviruses may share analogy to a certain extent to HCV.

More studies are now required to understand how these various proteins and non-protein components assemble to form the molecular machineries required for

polyprotein processing and viral RNA replication. As the newly generated RNA could be used for viral protein translation, viral RNA replication, or viral particle assembly, the decision making process between these options is still a black box. More systemic approaches at both the molecular and cellular level are needed in order to answer these important questions relating to the Flavivirus replication cycle. It is hopeful that these studies will promote the discovery of novel putative strategies for antiviral intervention.

3.3.3 Non-enzymatic roles played by NS3

For several (+) strand RNA viruses including poliovirus and flaviviruses, RNA replication, translation and encapsidation occur in a concerted manner, presumably to minimize amplification and packaging of defective RNA (Novak and Kirkegaard, 1994). Virus assembly occurs at the late state of viral infection. Although nonstructural proteins are not reported to be packed into progeny virions, they play important but less clearly defined roles for production of infectious virus (Murray *et al.*, 2008). NS3 plays a role in virus assembly, independent of its known enzymatic functions and in association with NS2A. Such a novel activity of NS3 has been demonstrated in YFV (Kummerer and Rice, 2002; Patkar and Kuhn, 2008). A second site changes in the helicase domain of NS3 (Asp 343 to uncharged residue) could restore infectious virus production in the absence of cleavage at the mutant NS2A α site. These findings suggest that physical association between the NS3 and NS2A are essential for infectious flavivirus morphogenesis. Apparently, NS3 protein must be supplied in *cis* for infectious virus production, as shown by complementation studies on KUNV replicon (Khromykh *et al.*, 2000; Liu *et al.*, 2002).

Given the strict conservation of the two bases AG at the 5' ends of both the + and – strand genomic flaviviral RNA, it is tempting to propose that having translocated towards the 5' end, the bound NS3 helicase would act as an “RNA anchoring device” by both preventing strand reannealing and helping redirect NS5 RdRp to the 3' end of the genome (Filomatori *et al.*, 2006), thereby ensuring the complete replication of RNA. Reasons for the asymmetry in genome replication and encapsidation might derive in part from specific recognition of capped viral RNA by the NS3 protease-helicase, presumably in association with NS2A (Patkar and Kuhn,

Chapter 3 Results and Discussion

2008) and capsid proteins. Clearly, NS3 is involved in almost all steps in viral polyprotein processing, genome replication and virus particle assembly, and thus constitutes a primary target for antiviral drug design. More research efforts are needed to elucidate how exactly the protein works for its different roles at different stages of viral life cycle.

Chapter 4 Conclusion and Future Directions

I have performed structural and functional studies on DENV4 NS2B-NS3 protease-helicase and NS3 helicase domain.

The non-structural protein 3 (NS3) from dengue virus carries out enzymatic reactions essential for viral replication. The serine protease N-terminal domain of NS3 with a segment of 40 residues from the NS2B protein acting as a cofactor is responsible for proteolysis of the polyprotein. The C-terminus of NS3 contains the ATPase/helicase domain. Atomic structures were available for these domains separately but a molecular view of the full length flavivirus NS3 polypeptide was still lacking. In order to study the overall architecture and its implications on the enzymatic activities of the NS2B-NS3 protease-helicase, a complete NS3 molecule fused to 18 residues of the NS2B cofactor through a flexible linker (scNS2B₁₈NS3) was engineered. The recombinant protein contains a protease domain that is enzymatically inactive but retains the same overall folding as the known active proteases from DENV2 and WNV. The protease domain is connected to the C terminal RNA helicase domain through a natural linker. The scNS2B₁₈NS3 protein was successfully expressed, purified and crystallized. The first conformation of scNS2B₁₈NS3 was determined at 3.15 Å resolution. The relative orientation between the protease and helicase domains is drastically different compared to the scNS3-NS4A molecule from hepatitis C virus (HCV), which was caught in the act of *cis* cleavage at the NS3-NS4A junction. Here, the protease domain sits beneath the ATP binding site, giving the molecule an elongated shape. The domain arrangement found in the crystal structure fits nicely into an envelope determined *ab-initio* using small angle X-ray scattering experiments in solution, suggesting a stable molecular conformation. The affinity for both ATP and ADP is approximately 10 fold higher for full length NS3 compared to the helicase domain, measured using fluorescence correlation spectroscopy. Thus, the protease domain plays an important regulatory role for NS3 NTPase and helicase activity. The second conformation of scNS2B₁₈NS3 was discovered during the process of improving protein quality. Although the two structures share very similar crystal form and both adopt an elongated shape, the protease domain in scNS2B₁₈NS3 Conformation II rotates ~ 160° relative to the

Chapter 4 Conclusion and Future Directions

helicase domain, in contrast to Conformation I. The structural dynamics of the two structures of scNS2B₁₈NS3 and the potential impacts to its enzymatic activity as well as its roles in viral life cycle are discussed.

In the second part of this work, the structure dynamics and the enzymatic activities of the NS3 helicase domain are studied. Together with the NS5 polymerase, the NS3 helicase plays a pivotal role in flavivirus RNA replication and constitutes an important drug target. The dengue virus NS3 helicase was captured at several stages along the catalytic pathway, namely, ligand-free state, binding to single stranded RNA, to an ATP analog, binding to a Vanadate-trapped ADP complex and finally binding to ATP hydrolysis products. RNA recognition appears largely sequence-independent in a way remarkably similar to eukaryotic DEAD box proteins Vasa and eIF4AIII. Extensive structural rearrangements of NS3 helicase upon ssRNA binding are observed: NS3 helicase switches to a catalytic-competent state imparted by an inward movement of the P-loop, interdomain closure and a change in the divalent metal coordination shell. The structural basis for RNA-stimulated ATP hydrolysis is provided based on these observations. In summary, these structures demonstrate for the first time large quaternary changes in the flavivirus helicase, identify the catalytic water molecule for ATP hydrolysis and point to a β -hairpin that protrudes from subdomain 2, as a critical element for dsRNA unwinding. They also suggest how NS3 could act as an RNA anchoring device and thus participate both in flavivirus RNA replication and assembly.

Future directions to unravel the molecular details of flavivirus replication machinery centering on the multifunctional NS3 protease-helicase are:

- To extend our current knowledge about the molecular mechanism of NS3 helicase activity: Based on the crystal structures of NS3 helicase ssRNA complex in this work, we proposed a model for the NS3 helicase translocation and unwinding mechanism. New structures of NS3 helicase in complex with dsRNA are of great interest. Intelligent design of short dsRNA substrates is very important for co-crystallization with NS3. Biochemical characterization of protein substrate interaction will assist the dsRNA substrate selection. Techniques like protein intrinsic

Chapter 4 Conclusion and Future Directions

fluorescence measurements, circular dichroism, isothermal titration calorimetry, Biacore (surface plasmon resonance), and NMR are good choices for such purpose.

- To further study the structural and functional properties of full length NS3 protein: The full length NS3 structure displays a dramatically different overall conformation from its homolog, Hepatitis C Virus NS3. Preliminary analysis also implies that certain structural flexibility exists and this flexibility plays important role in protein enzymatic activity regulation. Combinations of structure based mutagenesis, protein enzymatic activity study *in vitro* and X-ray crystallography can gain us more insights about this multi-functional protein.

- It is also interesting to further explore the potential of NS3 as a drug target. We will take advantage of our knowledge about NS3 structure and apply it to the adventure of searching for specific inhibitors. X-ray crystallography has become indispensable in the modern drug discovery journey. The near-atomic resolution 3D view of protein ligand interactions is helpful in inhibitor hit selection and Lead Optimization.

- To characterize the roles of NS3 in viral replication *in vitro*, i.e. its interaction with NS5 polymerase, viral genomic RNA and other cofactors in viral replication complex.

References

- Ahlquist, P., Noueiry, A.O., Lee, W.M., Kushner, D.B., and Dye, B.T. (2003). Host factors in positive-strand RNA virus genome replication. *J Virol* 77, 8181-8186.
- Aleshin, A.E., Shiryayev, S.A., Strongin, A.Y., and Liddington, R.C. (2007). Structural evidence for regulation and specificity of flaviviral proteases and evolution of the Flaviviridae fold. *Protein Sci* 16, 795-806.
- Ali, J.A., Maluf, N.K., and Lohman, T.M. (1999). An oligomeric form of E. coli UvrD is required for optimal helicase activity. *J Mol Biol* 293, 815-834.
- Allison, S.L., Stadler, K., Mandl, C.W., Kunz, C., and Heinz, F.X. (1995). Synthesis and secretion of recombinant tick-borne encephalitis virus protein E in soluble and particulate form. *J Virol* 69, 5816-5820.
- Alvarez, D.E., De Lella Ezcurra, A.L., Fucito, S., and Gamarnik, A.V. (2005a). Role of RNA structures present at the 3'UTR of dengue virus on translation, RNA synthesis, and viral replication. *Virology* 339, 200-212.
- Alvarez, D.E., Filomatori, C.V., and Gamarnik, A.V. (2008). Functional analysis of dengue virus cyclization sequences located at the 5' and 3'UTRs. *Virology* 375, 223-235.
- Alvarez, D.E., Lodeiro, M.F., Luduena, S.J., Pietrasanta, L.I., and Gamarnik, A.V. (2005b). Long-range RNA-RNA interactions circularize the dengue virus genome. *J Virol* 79, 6631-6643.
- Andersen, C.B., Ballut, L., Johansen, J.S., Chamieh, H., Nielsen, K.H., Oliveira, C.L., Pedersen, J.S., Seraphin, B., Le Hir, H., and Andersen, G.R. (2006). Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. *Science* 313, 1968-1972.
- Appel, N., Schaller, T., Penin, F., and Bartenschlager, R. (2006). From structure to function: new insights into hepatitis C virus RNA replication. *J Biol Chem* 281, 9833-9836.
- Arias, C.F., Preugschat, F., and Strauss, J.H. (1993). Dengue 2 virus NS2B and NS3 form a stable complex that can cleave NS3 within the helicase domain. *Virology* 193, 888-899.
- Armbruster, A., Svergun, D.I., Coskun, U., Juliano, S., Bailer, S.M., and Gruber, G. (2004). Structural analysis of the stalk subunit Vma5p of the yeast V-ATPase in solution. *FEBS Lett* 570, 119-125.
- Ashburn, P.M., and Craig, C.F. (2004). Experimental investigations regarding the etiology of dengue fever. 1907. *J Infect Dis* 189, 1747-1783; discussion 1744-1746.
- Assenberg, R., Ren, J., Verma, A., Walter, T.S., Alderton, D., Hurrelbrink, R.J., Fuller, S.D., Bressanelli, S., Owens, R.J., Stuart, D.I., and Grimes, J.M. (2007).

References

- Crystal structure of the Murray Valley encephalitis virus NS5 methyltransferase domain in complex with cap analogues. *J Gen Virol* 88, 2228-2236.
- Bacia, K., and Schwille, P. (2007). Fluorescence correlation spectroscopy. *Methods Mol Biol* 398, 73-84.
- Banerjee, R., and Dasgupta, A. (2001). Specific interaction of hepatitis C virus protease/helicase NS3 with the 3'-terminal sequences of viral positive- and negative-strand RNA. *J Virol* 75, 1708-1721.
- Bartelma, G., and Padmanabhan, R. (2002). Expression, purification, and characterization of the RNA 5'-triphosphatase activity of dengue virus type 2 nonstructural protein 3. *Virology* 299, 122-132.
- Bartholomeusz, A.I., and Wright, P.J. (1993). Synthesis of dengue virus RNA in vitro: initiation and the involvement of proteins NS3 and NS5. *Arch Virol* 128, 111-121.
- Benarroch, D., Selisko, B., Locatelli, G.A., Maga, G., Romette, J.L., and Canard, B. (2004). The RNA helicase, nucleotide 5'-triphosphatase, and RNA 5'-triphosphatase activities of Dengue virus protein NS3 are Mg²⁺-dependent and require a functional Walker B motif in the helicase catalytic core. *Virology* 328, 208-218.
- Benz, J., Trachsel, H., and Baumann, U. (1999). Crystal structure of the ATPase domain of translation initiation factor 4A from *Saccharomyces cerevisiae*--the prototype of the DEAD box protein family. *Structure* 7, 671-679.
- Bera, A.K., Kuhn, R.J., and Smith, J.L. (2007). Functional characterization of cis and trans activity of the Flavivirus NS2B-NS3 protease. *J Biol Chem* 282, 12883-12892.
- Beran, R.K., and Pyle, A.M. (2008). Hepatitis C viral NS3-4A protease activity is enhanced by the NS3 helicase. *J Biol Chem* 283, 29929-29937.
- Beran, R.K., Serebrov, V., and Pyle, A.M. (2007). The serine protease domain of hepatitis C viral NS3 activates RNA helicase activity by promoting the binding of RNA substrate. *J Biol Chem* 282, 34913-34920.
- Betterton, M.D., and Julicher, F. (2005). Opening of nucleic-acid double strands by helicases: active versus passive opening. *Phys Rev E Stat Nonlin Soft Matter Phys* 71, 011904.
- Blackwell, J.L., and Brinton, M.A. (1997). Translation elongation factor-1 alpha interacts with the 3' stem-loop region of West Nile virus genomic RNA. *J Virol* 71, 6433-6444.
- Bleichert, F., and Baserga, S.J. (2007). The long unwinding road of RNA helicases. *Mol Cell* 27, 339-352.
- Bono, F., Ebert, J., Lorentzen, E., and Conti, E. (2006). The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. *Cell* 126, 713-725.

References

- Borowski, P., Mueller, O., Niebuhr, A., Kalitzky, M., Hwang, L.H., Schmitz, H., Siwecka, M.A., and Kulikowski, T. (2000). ATP-binding domain of NTPase/helicase as a target for hepatitis C antiviral therapy. *Acta Biochim Pol* 47, 173-180.
- Borowski, P., Niebuhr, A., Schmitz, H., Hosmane, R.S., Bretner, M., Siwecka, M.A., and Kulikowski, T. (2002). NTPase/helicase of Flaviviridae: inhibitors and inhibition of the enzyme. *Acta Biochim Pol* 49, 597-614.
- Brass, V., Berke, J.M., Montserret, R., Blum, H.E., Penin, F., and Moradpour, D. (2008). Structural determinants for membrane association and dynamic organization of the hepatitis C virus NS3-4A complex. *Proc Natl Acad Sci U S A* 105, 14545-14550.
- Brooks, A.J., Johansson, M., John, A.V., Xu, Y., Jans, D.A., and Vasudevan, S.G. (2002). The interdomain region of dengue NS5 protein that binds to the viral helicase NS3 contains independently functional importin beta 1 and importin alpha/beta-recognized nuclear localization signals. *J Biol Chem* 277, 36399-36407.
- Buttner, K., Nehring, S., and Hopfner, K.P. (2007). Structural basis for DNA duplex separation by a superfamily-2 helicase. *Nat Struct Mol Biol* 14, 647-652.
- Cahour, A., Falgout, B., and Lai, C.J. (1992). 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* 66, 1535-1542.
- Calisher, C.H., Karabatsos, N., Dalrymple, J.M., Shope, R.E., Porterfield, J.S., Westaway, E.G., and Brandt, W.E. (1989). Antigenic relationships between flaviviruses as determined by cross-neutralization tests with polyclonal antisera. *J Gen Virol* 70 (Pt 1), 37-43.
- Caruthers, J.M., Johnson, E.R., and McKay, D.B. (2000). Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase. *Proc Natl Acad Sci U S A* 97, 13080-13085.
- Caruthers, J.M., and McKay, D.B. (2002). Helicase structure and mechanism. *Curr Opin Struct Biol* 12, 123-133.
- CCP4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50, 760-763.
- Chambers, T.J., Grakoui, A., and Rice, C.M. (1991). Processing of the yellow fever virus nonstructural polyprotein: a catalytically active NS3 proteinase domain and NS2B are required for cleavages at dibasic sites. *J Virol* 65, 6042-6050.
- Chambers, T.J., Hahn, C.S., Galler, R., and Rice, C.M. (1990a). Flavivirus genome organization, expression, and replication. *Annu Rev Microbiol* 44, 649-688.
- Chambers, T.J., McCourt, D.W., and Rice, C.M. (1990b). Production of yellow fever virus proteins in infected cells: identification of discrete polyprotein species and analysis of cleavage kinetics using region-specific polyclonal antisera. *Virology* 177, 159-174.

References

- Chambers, T.J., Nestorowicz, A., Amberg, S.M., and Rice, C.M. (1993). Mutagenesis of the yellow fever virus NS2B protein: effects on proteolytic processing, NS2B-NS3 complex formation, and viral replication. *J Virol* 67, 6797-6807.
- Chambers, T.J., Weir, R.C., Grakoui, A., McCourt, D.W., Bazan, J.F., Fletterick, R.J., and Rice, C.M. (1990c). 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* 87, 8898-8902.
- Chang, J.H., Cho, Y.H., Sohn, S.Y., Choi, J.M., Kim, A., Kim, Y.C., Jang, S.K., and Cho, Y. (2009). Crystal structure of the eIF4A-PDCD4 complex. *Proc Natl Acad Sci U S A*.
- Chen, C., Saxena, A.K., Simcoke, W.N., Garboczi, D.N., Pedersen, P.L., and Ko, Y.H. (2006). Mitochondrial ATP synthase. Crystal structure of the catalytic F1 unit in a vanadate-induced transition-like state and implications for mechanism. *J Biol Chem* 281, 13777-13783.
- Chen, C.J., Kuo, M.D., Chien, L.J., Hsu, S.L., Wang, Y.M., and Lin, J.H. (1997a). RNA-protein interactions: involvement of NS3, NS5, and 3' noncoding regions of Japanese encephalitis virus genomic RNA. *J Virol* 71, 3466-3473.
- Chen, Y., Maguire, T., Hileman, R.E., Fromm, J.R., Esko, J.D., Linhardt, R.J., and Marks, R.M. (1997b). Dengue virus infectivity depends on envelope protein binding to target cell heparan sulfate. *Nat Med* 3, 866-871.
- Chen, Y.C., Wang, S.Y., and King, C.C. (1999). Bacterial lipopolysaccharide inhibits dengue virus infection of primary human monocytes/macrophages by blockade of virus entry via a CD14-dependent mechanism. *J Virol* 73, 2650-2657.
- Cheng, W., Dumont, S., Tinoco, I., Jr., and Bustamante, C. (2007a). NS3 helicase actively separates RNA strands and senses sequence barriers ahead of the opening fork. *Proc Natl Acad Sci U S A* 104, 13954-13959.
- Cheng, Z., Collier, J., Parker, R., and Song, H. (2005). Crystal structure and functional analysis of DEAD-box protein Dhh1p. *RNA* 11, 1258-1270.
- Cheng, Z., Muhlrads, D., Lim, M.K., Parker, R., and Song, H. (2007b). Structural and functional insights into the human Upf1 helicase core. *EMBO J* 26, 253-264.
- Chernov, A.V., Shiryayev, S.A., Aleshin, A.E., Ratnikov, B.I., Smith, J.W., Liddington, R.C., and Strongin, A.Y. (2008). The two-component NS2B-NS3 proteinase represses DNA unwinding activity of the West Nile virus NS3 helicase. *J Biol Chem* 283, 17270-17278.
- Chitarra, V., Alzari, P.M., Bentley, G.A., Bhat, T.N., Eisele, J.L., Houdusse, A., Lescar, J., Souchon, H., and Poljak, R.J. (1993). Three-dimensional structure of a heteroclitic antigen-antibody cross-reaction complex. *Proc Natl Acad Sci U S A* 90, 7711-7715.

References

- Cho, H.S., Ha, N.C., Kang, L.W., Chung, K.M., Back, S.H., Jang, S.K., and Oh, B.H. (1998). Crystal structure of RNA helicase from genotype 1b hepatitis C virus. A feasible mechanism of unwinding duplex RNA. *J Biol Chem* 273, 15045-15052.
- Chu, P.W., and Westaway, E.G. (1985). Replication strategy of Kunjin virus: evidence for recycling role of replicative form RNA as template in semiconservative and asymmetric replication. *Virology* 140, 68-79.
- Chu, P.W., and Westaway, E.G. (1987). Characterization of Kunjin virus RNA-dependent RNA polymerase: reinitiation of synthesis in vitro. *Virology* 157, 330-337.
- Chung, K.M., Nybakken, G.E., Thompson, B.S., Engle, M.J., Marri, A., Fremont, D.H., and Diamond, M.S. (2006). Antibodies against West Nile Virus nonstructural protein NS1 prevent lethal infection through Fc gamma receptor-dependent and -independent mechanisms. *J Virol* 80, 1340-1351.
- Cleaves, G.R., and Dubin, D.T. (1979). Methylation status of intracellular dengue type 2 40 S RNA. *Virology* 96, 159-165.
- Clum, S., Ebner, K.E., and Padmanabhan, R. (1997). 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* 272, 30715-30723.
- Clyde, K., Barrera, J., and Harris, E. (2008). The capsid-coding region hairpin element (cHP) is a critical determinant of dengue virus and West Nile virus RNA synthesis. *Virology* 379, 314-323.
- Clyde, K., and Harris, E. (2006). 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* 80, 2170-2182.
- Clyde, K., Kyle, J.L., and Harris, E. (2006). Recent advances in deciphering viral and host determinants of dengue virus replication and pathogenesis. *J Virol* 80, 11418-11431.
- Collaborative, C.P.N. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50, 760-763.
- Cordin, O., Banroques, J., Tanner, N.K., and Linder, P. (2006). The DEAD-box protein family of RNA helicases. *Gene* 367, 17-37.
- Cordin, O., Tanner, N.K., Doere, M., Linder, P., and Banroques, J. (2004). The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity. *EMBO J* 23, 2478-2487.
- Cserzo, M., Wallin, E., Simon, I., von Heijne, G., and Elofsson, A. (1997). Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein Eng* 10, 673-676.

References

- Cui, T., Sugrue, R.J., Xu, Q., Lee, A.K., Chan, Y.C., and Fu, J. (1998). Recombinant dengue virus type 1 NS3 protein exhibits specific viral RNA binding and NTPase activity regulated by the NS5 protein. *Virology* 246, 409-417.
- De Nova-Ocampo, M., Villegas-Sepulveda, N., and del Angel, R.M. (2002). Translation elongation factor-1alpha, La, and PTB interact with the 3' untranslated region of dengue 4 virus RNA. *Virology* 295, 337-347.
- DeLano, W.L. (2002). The PyMOL User's Manual (DeLano Scientific, Palo Alto, CA, USA.).
- Denison, M.R. (2008). Seeking membranes: positive-strand RNA virus replication complexes. *PLoS Biol* 6, e270.
- Dessinges, M.N., Lionnet, T., Xi, X.G., Bensimon, D., and Croquette, V. (2004). Single-molecule assay reveals strand switching and enhanced processivity of UvrD. *Proc Natl Acad Sci U S A* 101, 6439-6444.
- Dimitrov, D.S. (2004). Virus entry: molecular mechanisms and biomedical applications. *Nat Rev Microbiol* 2, 109-122.
- Dong, H., Ray, D., Ren, S., Zhang, B., Puig-Basagoiti, F., Takagi, Y., Ho, C.K., Li, H., and Shi, P.Y. (2007). Distinct RNA elements confer specificity to flavivirus RNA cap methylation events. *J Virol* 81, 4412-4421.
- Dong, H., Zhang, B., and Shi, P.Y. (2008). Flavivirus methyltransferase: a novel antiviral target. *Antiviral Res* 80, 1-10.
- Edward, Z., and Takegami, T. (1993). Localization and functions of Japanese encephalitis virus nonstructural proteins NS3 and NS5 for viral RNA synthesis in the infected cells. *Microbiol Immunol* 37, 239-243.
- Egloff, M.P., Benarroch, D., Selisko, B., Romette, J.L., and Canard, B. (2002). An RNA cap (nucleoside-2'-O-)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization. *Embo J* 21, 2757-2768.
- Egloff, M.P., Decroly, E., Malet, H., Selisko, B., Benarroch, D., Ferron, F., and Canard, B. (2007). Structural and functional analysis of methylation and 5'-RNA sequence requirements of short capped RNAs by the methyltransferase domain of dengue virus NS5. *J Mol Biol* 372, 723-736.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60, 2126-2132.
- Enemark, E.J., and Joshua-Tor, L. (2008). On helicases and other motor proteins. *Curr Opin Struct Biol* 18, 243-257.
- Erbel, P., Schiering, N., D'Arcy, A., Renatus, M., Kroemer, M., Lim, S.P., Yin, Z., Keller, T.H., Vasudevan, S.G., and Hommel, U. (2006). Structural basis for the activation of flaviviral NS3 proteases from dengue and West Nile virus. *Nat Struct Mol Biol* 13, 372-373.

References

- Eryilmaz, J., Ceschini, S., Ryan, J., Geddes, S., Waters, T.R., and Barrett, T.E. (2006). Structural insights into the cryptic DNA-dependent ATPase activity of UvrB. *J Mol Biol* 357, 62-72.
- Falgout, B., Chanock, R., and Lai, C.J. (1989). Proper processing of dengue virus nonstructural glycoprotein NS1 requires the N-terminal hydrophobic signal sequence and the downstream nonstructural protein NS2a. *J Virol* 63, 1852-1860.
- Falgout, B., and Markoff, L. (1995). Evidence that flavivirus NS1-NS2A cleavage is mediated by a membrane-bound host protease in the endoplasmic reticulum. *J Virol* 69, 7232-7243.
- Falgout, B., Miller, R.H., and Lai, C.J. (1993). Deletion analysis of dengue virus type 4 nonstructural protein NS2B: identification of a domain required for NS2B-NS3 protease activity. *J Virol* 67, 2034-2042.
- Falgout, B., Pethel, M., Zhang, Y.M., and Lai, C.J. (1991). Both nonstructural proteins NS2B and NS3 are required for the proteolytic processing of dengue virus nonstructural proteins. *J Virol* 65, 2467-2475.
- Filomatori, C.V., Lodeiro, M.F., Alvarez, D.E., Samsa, M.M., Pietrasanta, L., and Gamarnik, A.V. (2006). A 5' RNA element promotes dengue virus RNA synthesis on a circular genome. *Genes Dev* 20, 2238-2249.
- Flamand, M., Megret, F., Mathieu, M., Lepault, J., Rey, F.A., and Deubel, V. (1999). Dengue virus type 1 nonstructural glycoprotein NS1 is secreted from mammalian cells as a soluble hexamer in a glycosylation-dependent fashion. *J Virol* 73, 6104-6110.
- Forwood, J.K., Brooks, A., Briggs, L.J., Xiao, C.Y., Jans, D.A., and Vasudevan, S.G. (1999). The 37-amino-acid interdomain of dengue virus NS5 protein contains a functional NLS and inhibitory CK2 site. *Biochem Biophys Res Commun* 257, 731-737.
- Frick, D.N. (2007). The hepatitis C virus NS3 protein: a model RNA helicase and potential drug target. *Curr Issues Mol Biol* 9, 1-20.
- Frick, D.N., Banik, S., and Rypma, R.S. (2007). Role of divalent metal cations in ATP hydrolysis catalyzed by the hepatitis C virus NS3 helicase: magnesium provides a bridge for ATP to fuel unwinding. *J Mol Biol* 365, 1017-1032.
- Frick, D.N., Rypma, R.S., Lam, A.M., and Frenz, C.M. (2004). Electrostatic analysis of the hepatitis C virus NS3 helicase reveals both active and allosteric site locations. *Nucleic Acids Res* 32, 5519-5528.
- Ganem, B. (1987). RNA world. *Nature* 328, 676.
- Germi, R., Crance, J.M., Garin, D., Guimet, J., Lortat-Jacob, H., Ruigrok, R.W., Zarski, J.P., and Drouet, E. (2002). Heparan sulfate-mediated binding of infectious dengue virus type 2 and yellow fever virus. *Virology* 292, 162-168.

References

- Gibbons, R.V., and Vaughn, D.W. (2002). Dengue: an escalating problem. *BMJ* 324, 1563-1566.
- Goncalvez, A.P., Escalante, A.A., Pujol, F.H., Ludert, J.E., Tovar, D., Salas, R.A., and Liprandi, F. (2002). Diversity and evolution of the envelope gene of dengue virus type 1. *Virology* 303, 110-119.
- Gorbalenya, A.E., and Koonin, E.V. (1993). Helicases: amino acid sequence comparisons and structure-function relationships. *Current opinion in structural biology*. 3, 419-429.
- Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P., and Blinov, V.M. (1988). 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* 235, 16-24.
- Grassmann, C.W., Isken, O., and Behrens, S.E. (1999). Assignment of the multifunctional NS3 protein of bovine viral diarrhea virus during RNA replication: an in vivo and in vitro study. *J Virol* 73, 9196-9205.
- Gross, C.H., and Shuman, S. (1996). Vaccinia virus RNA helicase: nucleic acid specificity in duplex unwinding. *J Virol* 70, 2615-2619.
- Gubler, D.J. (1998). Dengue and dengue hemorrhagic fever. *Clin Microbiol Rev* 11, 480-496.
- Gubler, D.J., Nalim, S., Tan, R., Saipan, H., and Sulianti Saroso, J. (1979). Variation in susceptibility to oral infection with dengue viruses among geographic strains of *Aedes aegypti*. *Am J Trop Med Hyg* 28, 1045-1052.
- Guinier, A., and Fournet, G. (1955). *Small angle scattering of X-rays* (Wiley, New York).
- Guzman, M.G., and Kouri, G. (2002). Dengue: an update. *Lancet Infect Dis* 2, 33-42.
- Guzman, M.G., Kouri, G., Valdes, L., Bravo, J., Vazquez, S., and Halstead, S.B. (2002). Enhanced severity of secondary dengue-2 infections: death rates in 1981 and 1997 Cuban outbreaks. *Rev Panam Salud Publica* 11, 223-227.
- Hahn, C.S., Hahn, Y.S., Rice, C.M., Lee, E., Dalgarno, L., Strauss, E.G., and Strauss, J.H. (1987). Conserved elements in the 3' untranslated region of flavivirus RNAs and potential cyclization sequences. *J Mol Biol* 198, 33-41.
- Hall, M.C., and Matson, S.W. (1999). Helicase motifs: the engine that powers DNA unwinding. *Mol Microbiol* 34, 867-877.
- Halstead, S.B. (1979). In vivo enhancement of dengue virus infection in rhesus monkeys by passively transferred antibody. *J Infect Dis* 140, 527-533.
- Halstead, S.B. (1983). Dengue haemorrhagic fever. *Trans R Soc Trop Med Hyg* 77, 739-740.

References

- Halstead, S.B. (1988). Pathogenesis of dengue: challenges to molecular biology. *Science* 239, 476-481.
- Halstead, S.B., Heinz, F.X., Barrett, A.D., and Roehrig, J.T. (2005). Dengue virus: molecular basis of cell entry and pathogenesis, 25-27 June 2003, Vienna, Austria. *Vaccine* 23, 849-856.
- Halstead, S.B., Rojanasuphot, S., and Sangkawibha, N. (1983). Original antigenic sin in dengue. *Am J Trop Med Hyg* 32, 154-156.
- Hayward, S., and Lee, R.A. (2002). Improvements in the analysis of domain motions in proteins from conformational change: DynDom version 1.50. *J Mol Graph Model* 21, 181-183.
- Heinz, F.X., MS, C., Purcell, R.H., Gould, E.A., Howard, C.R., Houghton, M., Moormann, R.J.M., Rice, C.M., and Thiel, H.J. (2000). Family flaviviridae, Vol Virus Taxonomy: 7th International Committee for the Taxonomy of Viruses (Academic Press, San Diego).
- Henchal, E.A., Henchal, L.S., and Schlesinger, J.J. (1988). Synergistic interactions of anti-NS1 monoclonal antibodies protect passively immunized mice from lethal challenge with dengue 2 virus. *J Gen Virol* 69 (Pt 8), 2101-2107.
- Henchal, E.A., and Putnak, J.R. (1990). The dengue viruses. *Clin Microbiol Rev* 3, 376-396.
- Hirokawa, T., Boon-Chieng, S., and Mitaku, S. (1998). SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* 14, 378-379.
- Hogbom, M., Collins, R., van den Berg, S., Jenvert, R.M., Karlberg, T., Kotenyova, T., Flores, A., Karlsson Hedestam, G.B., and Schiavone, L.H. (2007). Crystal structure of conserved domains 1 and 2 of the human DEAD-box helicase DDX3X in complex with the mononucleotide AMP. *J Mol Biol* 372, 150-159.
- Holmes, E.C., and Twiddy, S.S. (2003). The origin, emergence and evolutionary genetics of dengue virus. *Infect Genet Evol* 3, 19-28.
- Hotta, S. (1952a). Experimental studies on dengue. I. Isolation, identification and modification of the virus. *J Infect Dis* 90, 1-9.
- Hotta, S. (1952b). Experimental studies on dengue. II. A skin reaction observed during the epidemic of Osaka, Japan, in 1944. *J Infect Dis* 90, 10-12.
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J Mol Graph* 14, 33-38, 27-38.
- Hunke, C., Chen, W.J., Schafer, H.J., and Gruber, G. (2007). Cloning, purification, and nucleotide-binding traits of the catalytic subunit A of the VIVO ATPase from *Aedes albopictus*. *Protein Expr Purif* 53, 378-383.

References

- Jankowsky, E., and Fairman, M.E. (2007). RNA helicases--one fold for many functions. *Curr Opin Struct Biol* 17, 316-324.
- Jankowsky, E., Gross, C.H., Shuman, S., and Pyle, A.M. (2000). The DEXH protein NPH-II is a processive and directional motor for unwinding RNA. *Nature* 403, 447-451.
- Jennings, T.A., Mackintosh, S.G., Harrison, M.K., Sikora, D., Sikora, B., Tackett, A.J., Cameron, C.E., and Raney, K.D. (2008). NS3 helicase from the hepatitis C Virus can function as a monomer or oligomer depending on enzyme and substrate concentrations. *J Biol Chem*.
- Jindadamrongwech, S., Thepparit, C., and Smith, D.R. (2004). Identification of GRP 78 (BiP) as a liver cell expressed receptor element for dengue virus serotype 2. *Arch Virol* 149, 915-927.
- Johansson, M., Brooks, A.J., Jans, D.A., and Vasudevan, S.G. (2001). A small region of the dengue virus-encoded RNA-dependent RNA polymerase, NS5, confers interaction with both the nuclear transport receptor importin-beta and the viral helicase, NS3. *J Gen Virol* 82, 735-745.
- Jones, S., and Thornton, J.M. (1996). Principles of protein-protein interactions. *Proc Natl Acad Sci U S A* 93, 13-20.
- Kapoor, M., Zhang, L., Ramachandra, M., Kusukawa, J., Ebner, K.E., and Padmanabhan, R. (1995). Association between NS3 and NS5 proteins of dengue virus type 2 in the putative RNA replicase is linked to differential phosphorylation of NS5. *J Biol Chem* 270, 19100-19106.
- Keller, T.H., Chen, Y.L., Knox, J.E., Lim, S.P., Ma, N.L., Patel, S.J., Sampath, A., Wang, Q.Y., Yin, Z., and Vasudevan, S.G. (2006). Finding new medicines for flaviviral targets. *Novartis Found Symp* 277, 102-114; discussion 114-109, 251-103.
- Khromykh, A.A., Kondratieva, N., Sgro, J.Y., Palmenberg, A., and Westaway, E.G. (2003). Significance in replication of the terminal nucleotides of the flavivirus genome. *J Virol* 77, 10623-10629.
- Khromykh, A.A., Meka, H., Guyatt, K.J., and Westaway, E.G. (2001a). Essential role of cyclization sequences in flavivirus RNA replication. *J Virol* 75, 6719-6728.
- Khromykh, A.A., Sedlak, P.L., and Westaway, E.G. (2000). cis- and trans-acting elements in flavivirus RNA replication. *J Virol* 74, 3253-3263.
- Khromykh, A.A., Varnavski, A.N., Sedlak, P.L., and Westaway, E.G. (2001b). Coupling between replication and packaging of flavivirus RNA: evidence derived from the use of DNA-based full-length cDNA clones of Kunjin virus. *J Virol* 75, 4633-4640.
- Khromykh, A.A., Varnavski, A.N., and Westaway, E.G. (1998). Encapsidation of the flavivirus kunjin replicon RNA by using a complementation system providing Kunjin virus structural proteins in trans. *J Virol* 72, 5967-5977.

References

- Kim, D.W., Gwack, Y., Han, J.H., and Choe, J. (1995). C-terminal domain of the hepatitis C virus NS3 protein contains an RNA helicase activity. *Biochem Biophys Res Commun* 215, 160-166.
- Kim, D.W., Gwack, Y., Han, J.H., and Choe, J. (1997). Towards defining a minimal functional domain for NTPase and RNA helicase activities of the hepatitis C virus NS3 protein. *Virus Res* 49, 17-25.
- Kim, J.L., Morgenstern, K.A., Griffith, J.P., Dwyer, M.D., Thomson, J.A., Murcko, M.A., Lin, C., and Caron, P.R. (1998). Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding. *Structure* 6, 89-100.
- Kim, J.L., Morgenstern, K.A., Lin, C., Fox, T., Dwyer, M.D., Landro, J.A., Chambers, S.P., Markland, W., Lepre, C.A., O'Malley, E.T., *et al.* (1996). Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide. *Cell* 87, 343-355.
- Kim, S.H., Smith, J., Claude, A., and Lin, R.J. (1992). The purified yeast pre-mRNA splicing factor PRP2 is an RNA-dependent NTPase. *EMBO J* 11, 2319-2326.
- Kohlstaedt, L.A., Wang, J., Friedman, J.M., Rice, P.A., and Steitz, T.A. (1992). Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* 256, 1783-1790.
- Konarev, P.V., Petoukhov, M.V., and Svergun, D.I. (2001). MASSHA - a graphics system for rigid-body modelling of macromolecular complexes against solution scattering data. *J. Appl. Cryst.* 34, 527-532.
- Koonin, E.V. (1991). The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *J Gen Virol* 72 (Pt 9), 2197-2206.
- Koonin, E.V., and Dolja, V.V. (1993). Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. *Crit Rev Biochem Mol Biol* 28, 375-430.
- Korolev, S., Hsieh, J., Gauss, G.H., Lohman, T.M., and Waksman, G. (1997). Major domain swiveling revealed by the crystal structures of complexes of *E. coli* Rep helicase bound to single-stranded DNA and ADP. *Cell* 90, 635-647.
- Korolev, S., Yao, N., Lohman, T.M., Weber, P.C., and Waksman, G. (1998). Comparisons between the structures of HCV and Rep helicases reveal structural similarities between SF1 and SF2 super-families of helicases. *Protein Sci* 7, 605-610.
- Kroschewski, H., Allison, S.L., Heinz, F.X., and Mandl, C.W. (2003). Role of heparan sulfate for attachment and entry of tick-borne encephalitis virus. *Virology* 308, 92-100.
- Kuhn, R.J., Zhang, W., Rossmann, M.G., Pletnev, S.V., Corver, J., Lenches, E., Jones, C.T., Mukhopadhyay, S., Chipman, P.R., Strauss, E.G., *et al.* (2002). Structure of dengue virus: implications for flavivirus organization, maturation, and fusion. *Cell* 108, 717-725.

References

- Kummerer, B.M., and Rice, C.M. (2002). Mutations in the yellow fever virus nonstructural protein NS2A selectively block production of infectious particles. *J Virol* 76, 4773-4784.
- Kwong, A.D., Kim, J.L., and Lin, C. (2000). Structure and function of hepatitis C virus NS3 helicase. *Curr Top Microbiol Immunol* 242, 171-196.
- Lam, A.M., Rypma, R.S., and Frick, D.N. (2004). Enhanced nucleic acid binding to ATP-bound hepatitis C virus NS3 helicase at low pH activates RNA unwinding. *Nucleic Acids Res* 32, 4060-4070.
- Laskowski, R.A., Macarthur M. W., Moss D. S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 26, 283-291.
- Lee, E., and Lobigs, M. (2002). Mechanism of virulence attenuation of glycosaminoglycan-binding variants of Japanese encephalitis virus and Murray Valley encephalitis virus. *J Virol* 76, 4901-4911.
- Lee, J.M., Crooks, A.J., and Stephenson, J.R. (1989). The synthesis and maturation of a non-structural extracellular antigen from tick-borne encephalitis virus and its relationship to the intracellular NS1 protein. *J Gen Virol* 70 (Pt 2), 335-343.
- Lee, J.Y., and Yang, W. (2006). UvrD helicase unwinds DNA one base pair at a time by a two-part power stroke. *Cell* 127, 1349-1360.
- Lescar, J., Luo, D., Xu, T., Sampath, A., Lim, S.P., Canard, B., and Vasudevan, S.G. (2008). Towards the design of antiviral inhibitors against flaviviruses: the case for the multifunctional NS3 protein from Dengue virus as a target. *Antiviral Res* 80, 94-101.
- Leung, D., Schroder, K., White, H., Fang, N.X., Stoermer, M.J., Abbenante, G., Martin, J.L., Young, P.R., and Fairlie, D.P. (2001). Activity of recombinant dengue 2 virus NS3 protease in the presence of a truncated NS2B co-factor, small peptide substrates, and inhibitors. *J Biol Chem* 276, 45762-45771.
- Leung, J.Y., Pijlman, G.P., Kondratieva, N., Hyde, J., Mackenzie, J.M., and Khromykh, A.A. (2008). Role of nonstructural protein NS2A in flavivirus assembly. *J Virol* 82, 4731-4741.
- Levin, M.K., Gurjar, M., and Patel, S.S. (2005). A Brownian motor mechanism of translocation and strand separation by hepatitis C virus helicase. *Nat Struct Mol Biol* 12, 429-435.
- Levin, M.K., Wang, Y.H., and Patel, S.S. (2004). The functional interaction of the hepatitis C virus helicase molecules is responsible for unwinding processivity. *J Biol Chem* 279, 26005-26012.
- Li, H., Clum, S., You, S., Ebner, K.E., and Padmanabhan, R. (1999). The serine protease and RNA-stimulated nucleoside triphosphatase and RNA helicase functional domains of dengue virus type 2 NS3 converge within a region of 20 amino acids. *J Virol* 73, 3108-3116.

References

- Li, J., Lim, S.P., Beer, D., Patel, V., Wen, D., Tumanut, C., Tully, D.C., Williams, J.A., Jiricek, J., Priestle, J.P., *et al.* (2005). Functional profiling of recombinant NS3 proteases from all four serotypes of dengue virus using tetrapeptide and octapeptide substrate libraries. *J Biol Chem* 280, 28766-28774.
- Li, L., Lok, S.M., Yu, I.M., Zhang, Y., Kuhn, R.J., Chen, J., and Rossmann, M.G. (2008). The flavivirus precursor membrane-envelope protein complex: structure and maturation. *Science* 319, 1830-1834.
- Lin, C., Amberg, S.M., Chambers, T.J., and Rice, C.M. (1993a). 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* 67, 2327-2335.
- Lin, C., Chambers, T.J., and Rice, C.M. (1993b). 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* 192, 596-604.
- Lin, C.F., Chiu, S.C., Hsiao, Y.L., Wan, S.W., Lei, H.Y., Shiao, A.L., Liu, H.S., Yeh, T.M., Chen, S.H., Liu, C.C., and Lin, Y.S. (2005). Expression of cytokine, chemokine, and adhesion molecules during endothelial cell activation induced by antibodies against dengue virus nonstructural protein 1. *J Immunol* 174, 395-403.
- Lindenbach, B.D., and Rice, C.M. (1997). trans-Complementation of yellow fever virus NS1 reveals a role in early RNA replication. *J Virol* 71, 9608-9617.
- Lindenbach, B.D., and Rice, C.M. (1999). Genetic interaction of flavivirus nonstructural proteins NS1 and NS4A as a determinant of replicase function. *J Virol* 73, 4611-4621.
- Lindenbach, B.D., Thiel, H.J., and Rice, C.M. (2007). *Flaviviridae: the viruses and their replication*. (Philadelphia: Lippincott-Raven Publishers.).
- Liu, W.J., Chen, H.B., and Khromykh, A.A. (2003). Molecular and functional analyses of Kunjin virus infectious cDNA clones demonstrate the essential roles for NS2A in virus assembly and for a nonconservative residue in NS3 in RNA replication. *J Virol* 77, 7804-7813.
- Liu, W.J., Sedlak, P.L., Kondratieva, N., and Khromykh, A.A. (2002). Complementation analysis of the flavivirus Kunjin NS3 and NS5 proteins defines the minimal regions essential for formation of a replication complex and shows a requirement of NS3 in cis for virus assembly. *J Virol* 76, 10766-10775.
- Lo, M.K., Tilgner, M., Bernard, K.A., and Shi, P.Y. (2003). Functional analysis of mosquito-borne flavivirus conserved sequence elements within 3' untranslated region of West Nile virus by use of a reporting replicon that differentiates between viral translation and RNA replication. *J Virol* 77, 10004-10014.
- Lobigs, M. (1992). Proteolytic processing of a Murray Valley encephalitis virus non-structural polyprotein segment containing the viral proteinase: accumulation of a NS3-4A precursor which requires mature NS3 for efficient processing. *J Gen Virol* 73 (Pt 9), 2305-2312.

References

- Locatelli, G.A., Spadari, S., and Maga, G. (2002). Hepatitis C virus NS3 ATPase/helicase: an ATP switch regulates the cooperativity among the different substrate binding sites. *Biochemistry* 41, 10332-10342.
- Loh, P.G., Yang, H.S., Walsh, M.A., Wang, Q., Wang, X., Cheng, Z., Liu, D., and Song, H. (2009). Structural basis for translational inhibition by the tumour suppressor Pdc4. *EMBO J* 28, 274-285.
- Love, R.A., Parge, H.E., Wickersham, J.A., Hostomsky, Z., Habuka, N., Moomaw, E.W., Adachi, T., and Hostomska, Z. (1996). The crystal structure of hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site. *Cell* 87, 331-342.
- Luo, D., Xu, T., Hunke, C., Gruber, G., Vasudevan, S.G., and Lescar, J. (2008a). Crystal structure of the NS3 protease-helicase from dengue virus. *J Virol* 82, 173-183.
- Luo, D., Xu, T., Watson, R.P., Scherer-Becker, D., Sampath, A., Jahnke, W., Yeong, S.S., Wang, C.H., Lim, S.P., Strongin, A., *et al.* (2008b). Insights into RNA unwinding and ATP hydrolysis by the flavivirus NS3 protein. *EMBO J* 27, 3209-3219.
- Mackenzie, J.M., Khromykh, A.A., Jones, M.K., and Westaway, E.G. (1998). Subcellular localization and some biochemical properties of the flavivirus Kunjin nonstructural proteins NS2A and NS4A. *Virology* 245, 203-215.
- Mackenzie, J.M., and Westaway, E.G. (2001). Assembly and maturation of the flavivirus Kunjin virus appear to occur in the rough endoplasmic reticulum and along the secretory pathway, respectively. *J Virol* 75, 10787-10799.
- Mackintosh, S.G., Lu, J.Z., Jordan, J.B., Harrison, M.K., Sikora, B., Sharma, S.D., Cameron, C.E., Raney, K.D., and Sakon, J. (2006). Structural and biological identification of residues on the surface of NS3 helicase required for optimal replication of the hepatitis C virus. *J Biol Chem* 281, 3528-3535.
- Malet, H., Egloff, M.P., Selisko, B., Butcher, R.E., Wright, P.J., Roberts, M., Gruez, A., Sulzenbacher, G., Vornrhein, C., Bricogne, G., *et al.* (2007). Crystal structure of the RNA polymerase domain of the West Nile virus non-structural protein 5. *J Biol Chem* 282, 10678-10689.
- Mancini, E.J., Assenberg, R., Verma, A., Walter, T.S., Tuma, R., Grimes, J.M., Owens, R.J., and Stuart, D.I. (2007). Structure of the Murray Valley encephalitis virus RNA helicase at 1.9 Angstrom resolution. *Protein Sci* 16, 2294-2300.
- Markoff, L. (2003). 5'- and 3'-noncoding regions in flavivirus RNA. *Adv Virus Res* 59, 177-228.
- Markoff, L., Pang, X., Hsueh, H.S., Falgout, B., Olsen, R., Jones, E., and Polo, S. (2002). Derivation and characterization of a dengue type 1 host range-restricted mutant virus that is attenuated and highly immunogenic in monkeys. *J Virol* 76, 3318-3328.
- Mason, P.W. (1989). Maturation of Japanese encephalitis virus glycoproteins produced by infected mammalian and mosquito cells. *Virology* 169, 354-364.

References

- Mastrangelo, E., Bollati, M., Milani, M., Brisbarre, N., de Lamballerie, X., Coutard, B., Canard, B., Khromykh, A., and Bolognesi, M. (2006). Preliminary crystallographic characterization of an RNA helicase from Kunjin virus. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 62, 876-879.
- Mastrangelo, E., Bollati, M., Milani, M., Selisko, B., Peyrane, F., Canard, B., Grard, G., de Lamballerie, X., and Bolognesi, M. (2007a). Structural bases for substrate recognition and activity in Meaban virus nucleoside-2'-O-methyltransferase. *Protein Sci* 16, 1133-1145.
- Mastrangelo, E., Milani, M., Bollati, M., Selisko, B., Peyrane, F., Pandini, V., Sorrentino, G., Canard, B., Konarev, P.V., Svergun, D.I., *et al.* (2007b). Crystal Structure and Activity of Kunjin Virus NS3 Helicase; Protease and Helicase Domain Assembly in the Full Length NS3 Protein. *J Mol Biol*.
- Mastrangelo, E., Milani, M., Bollati, M., Selisko, B., Peyrane, F., Pandini, V., Sorrentino, G., Canard, B., Konarev, P.V., Svergun, D.I., *et al.* (2007c). Crystal structure and activity of Kunjin virus NS3 helicase; protease and helicase domain assembly in the full length NS3 protein. *J Mol Biol* 372, 444-455.
- Matthews, B.W. (1968). Solvent content of protein crystals. *J Mol Biol* 33, 491-497.
- Matusan, A.E., Pryor, M.J., Davidson, A.D., and Wright, P.J. (2001). Mutagenesis of the Dengue virus type 2 NS3 protein within and outside helicase motifs: effects on enzyme activity and virus replication. *J Virol* 75, 9633-9643.
- McCoy, A.J. (2007). Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr D Biol Crystallogr* 63, 32-41.
- Meyers, B.C., Matzke, M., and Sundaresan, V. (2008). The RNA world is alive and well. *Trends Plant Sci* 13, 311-313.
- Miller, S., Kastner, S., Krijnse-Locker, J., Buhler, S., and Bartenschlager, R. (2007). The non-structural protein 4A of dengue virus is an integral membrane protein inducing membrane alterations in a 2K-regulated manner. *J Biol Chem* 282, 8873-8882.
- Miller, S., Sparacio, S., and Bartenschlager, R. (2006). Subcellular localization and membrane topology of the Dengue virus type 2 Non-structural protein 4B. *J Biol Chem* 281, 8854-8863.
- Modis, Y., Ogata, S., Clements, D., and Harrison, S.C. (2004). Structure of the dengue virus envelope protein after membrane fusion. *Nature* 427, 313-319.
- Mukhopadhyay, S., Kuhn, R.J., and Rossmann, M.G. (2005). A structural perspective of the flavivirus life cycle. *Nat Rev Microbiol* 3, 13-22.
- Muller, J.D., Chen, Y., and Gratton, E. (2003). Fluorescence correlation spectroscopy. *Methods Enzymol* 361, 69-92.

References

- Murray, C.L., Jones, C.T., and Rice, C.M. (2008). Architects of assembly: roles of Flaviviridae non-structural proteins in virion morphogenesis. *Nat Rev Microbiol* 6, 699-708.
- Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53, 240-255.
- Murthy, H.M., Clum, S., and Padmanabhan, R. (1999). Dengue virus NS3 serine protease. Crystal structure and insights into interaction of the active site with substrates by molecular modeling and structural analysis of mutational effects. *J Biol Chem* 274, 5573-5580.
- Muylaert, I.R., Galler, R., and Rice, C.M. (1997). Genetic analysis of the yellow fever virus NS1 protein: identification of a temperature-sensitive mutation which blocks RNA accumulation. *J Virol* 71, 291-298.
- Myong, S., Bruno, M.M., Pyle, A.M., and Ha, T. (2007). Spring-loaded mechanism of DNA unwinding by hepatitis C virus NS3 helicase. *Science* 317, 513-516.
- Navarro-Sanchez, E., Altmeyer, R., Amara, A., Schwartz, O., Fieschi, F., Virelizier, J.L., Arenzana-Seisdedos, F., and Despres, P. (2003). Dendritic-cell-specific ICAM3-grabbing non-integrin is essential for the productive infection of human dendritic cells by mosquito-cell-derived dengue viruses. *EMBO Rep* 4, 723-728.
- Nestorowicz, A., Chambers, T.J., and Rice, C.M. (1994). 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* 199, 114-123.
- Niedziela-Majka, A., Chesnik, M.A., Tomko, E.J., and Lohman, T.M. (2007). *Bacillus stearothermophilus* PcrA monomer is a single-stranded DNA translocase but not a processive helicase in vitro. *J Biol Chem* 282, 27076-27085.
- Nielsen, K.H., Chamieh, H., Andersen, C.B., Fredslund, F., Hamborg, K., Le Hir, H., and Andersen, G.R. (2009). Mechanism of ATP turnover inhibition in the EJC. *RNA* 15, 67-75.
- Nomaguchi, M., Ackermann, M., Yon, C., You, S., and Padmanabhan, R. (2003). De novo synthesis of negative-strand RNA by Dengue virus RNA-dependent RNA polymerase in vitro: nucleotide, primer, and template parameters. *J Virol* 77, 8831-8842.
- North, G. (1987). Back to the RNA world--and beyond. *Nature* 328, 18-19.
- Novak, J.E., and Kirkegaard, K. (1994). Coupling between genome translation and replication in an RNA virus. *Genes Dev* 8, 1726-1737.
- Painter, J., and Merritt, E.A. (2006). Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr* 62, 439-450.

References

- Pang, P.S., Jankowsky, E., Planet, P.J., and Pyle, A.M. (2002). The hepatitis C viral NS3 protein is a processive DNA helicase with cofactor enhanced RNA unwinding. *EMBO J* 21, 1168-1176.
- Parrish, C.R., Woo, W.S., and Wright, P.J. (1991). Expression of the NS1 gene of dengue virus type 2 using vaccinia virus. Dimerisation of the NS1 glycoprotein. *Arch Virol* 117, 279-286.
- Patkar, C.G., and Kuhn, R.J. (2008). Yellow Fever virus NS3 plays an essential role in virus assembly independent of its known enzymatic functions. *J Virol* 82, 3342-3352.
- Pause, A., Methot, N., and Sonenberg, N. (1993). The HRIGRXXR region of the DEAD box RNA helicase eukaryotic translation initiation factor 4A is required for RNA binding and ATP hydrolysis. *Mol Cell Biol* 13, 6789-6798.
- Pause, A., and Sonenberg, N. (1992). Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. *EMBO J* 11, 2643-2654.
- Polacek, C., Foley, J.E., and Harris, E. (2009). Conformational changes in the solution structure of the dengue virus 5' end in the presence and absence of the 3' untranslated region. *J Virol* 83, 1161-1166.
- Preugschat, F., Averett, D.R., Clarke, B.E., and Porter, D.J. (1996). A steady-state and pre-steady-state kinetic analysis of the NTPase activity associated with the hepatitis C virus NS3 helicase domain. *J Biol Chem* 271, 24449-24457.
- Preugschat, F., Yao, C.W., and Strauss, J.H. (1990). In vitro processing of dengue virus type 2 nonstructural proteins NS2A, NS2B, and NS3. *J Virol* 64, 4364-4374.
- Pryor, M.J., and Wright, P.J. (1994). Glycosylation mutants of dengue virus NS1 protein. *J Gen Virol* 75 (Pt 5), 1183-1187.
- Pyle, A.M. (2008). Translocation and Unwinding Mechanisms of RNA and DNA Helicases. *Annual Review of Biophysics* 37, 317-336.
- Rajkowitsch, L., Chen, D., Stampfl, S., Semrad, K., Waldsich, C., Mayer, O., Jantsch, M.F., Konrat, R., Blasi, U., and Schroeder, R. (2007). RNA chaperones, RNA annealers and RNA helicases. *RNA Biol* 4, 118-130.
- Rawlinson, S.M., Pryor, M.J., Wright, P.J., and Jans, D.A. (2006). Dengue virus RNA polymerase NS5: a potential therapeutic target? *Curr Drug Targets* 7, 1623-1638.
- Ray, D., Shah, A., Tilgner, M., Guo, Y., Zhao, Y., Dong, H., Deas, T.S., Zhou, Y., Li, H., and Shi, P.Y. (2006). West Nile virus 5'-cap structure is formed by sequential guanine N-7 and ribose 2'-O methylations by nonstructural protein 5. *J Virol* 80, 8362-8370.
- Reed, W., and Carroll, J. (2001). The etiology of yellow fever: a supplemental note. 1902. *Mil Med* 166, 62-66.

References

- Rey, F.A., Heinz, F.X., Mandl, C., Kunz, C., and Harrison, S.C. (1995). The envelope glycoprotein from tick-borne encephalitis virus at 2 Å resolution. *Nature* 375, 291-298.
- Rice, C.M., Lenches, E.M., Eddy, S.R., Shin, S.J., Sheets, R.L., and Strauss, J.H. (1985). Nucleotide sequence of yellow fever virus: implications for flavivirus gene expression and evolution. *Science* 229, 726-733.
- Rico-Hesse, R. (1990). Molecular evolution and distribution of dengue viruses type 1 and 2 in nature. *Virology* 174, 479-493.
- Rocak, S., Emery, B., Tanner, N.K., and Linder, P. (2005). Characterization of the ATPase and unwinding activities of the yeast DEAD-box protein Has1p and the analysis of the roles of the conserved motifs. *Nucleic Acids Res* 33, 999-1009.
- Rocak, S., and Linder, P. (2004). DEAD-box proteins: the driving forces behind RNA metabolism. *Nat Rev Mol Cell Biol* 5, 232-241.
- Rogers, G.W., Jr., Komar, A.A., and Merrick, W.C. (2002). eIF4A: the godfather of the DEAD box helicases. *Prog Nucleic Acid Res Mol Biol* 72, 307-331.
- Roosendaal, J., Westaway, E.G., Khromykh, A., and Mackenzie, J.M. (2006). 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* 80, 4623-4632.
- Sabin, A.B. (1952). Research on dengue during World War II. *Am J Trop Med Hyg* 1, 30-50.
- Salonen, A., Ahola, T., and Kaariainen, L. (2005). Viral RNA replication in association with cellular membranes. *Curr Top Microbiol Immunol* 285, 139-173.
- Sampath, A., Xu, T., Chao, A., Luo, D., Lescar, J., and Vasudevan, S.G. (2006). Structure-based mutational analysis of the NS3 helicase from dengue virus. *J Virol* 80, 6686-6690.
- Schutz, P., Bumann, M., Oberholzer, A.E., Bieniossek, C., Trachsel, H., Altmann, M., and Baumann, U. (2008). Crystal structure of the yeast eIF4A-eIF4G complex: an RNA-helicase controlled by protein-protein interactions. *Proc Natl Acad Sci U S A* 105, 9564-9569.
- Schwer, B., and Meszaros, T. (2000). RNA helicase dynamics in pre-mRNA splicing. *EMBO J* 19, 6582-6591.
- Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S., and Yokoyama, S. (2006). Structural basis for RNA unwinding by the DEAD-box protein Drosophila Vasa. *Cell* 125, 287-300.
- Shi, H., Cordin, O., Minder, C.M., Linder, P., and Xu, R.M. (2004). Crystal structure of the human ATP-dependent splicing and export factor UAP56. *Proc Natl Acad Sci U S A* 101, 17628-17633.

References

- Silverman, E., Edwalds-Gilbert, G., and Lin, R.J. (2003). DExD/H-box proteins and their partners: helping RNA helicases unwind. *Gene* 312, 1-16.
- Singleton, M.R., Dillingham, M.S., and Wigley, D.B. (2007). Structure and mechanism of helicases and nucleic acid translocases. *Annu Rev Biochem* 76, 23-50.
- Smith, C.A., and Rayment, I. (1996). X-ray structure of the magnesium(II).ADP.vanadate complex of the Dictyostelium discoideum myosin motor domain to 1.9 Å resolution. *Biochemistry* 35, 5404-5417.
- Soultanas, P., Dillingham, M.S., Velankar, S.S., and Wigley, D.B. (1999). DNA binding mediates conformational changes and metal ion coordination in the active site of PcrA helicase. *J Mol Biol* 290, 137-148.
- Speroni, S., De Colibus, L., Mastrangelo, E., Gould, E., Coutard, B., Forrester, N.L., Blanc, S., Canard, B., and Mattevi, A. (2008). Structure and biochemical analysis of Kokobera virus helicase. *Proteins* 70, 1120-1123.
- Stadler, K., Allison, S.L., Schalich, J., and Heinz, F.X. (1997). Proteolytic activation of tick-borne encephalitis virus by furin. *J Virol* 71, 8475-8481.
- Stiasny, K., Allison, S.L., Mandl, C.W., and Heinz, F.X. (2001). Role of metastability and acidic pH in membrane fusion by tick-borne encephalitis virus. *J Virol* 75, 7392-7398.
- Stiasny, K., Allison, S.L., Marchler-Bauer, A., Kunz, C., and Heinz, F.X. (1996). Structural requirements for low-pH-induced rearrangements in the envelope glycoprotein of tick-borne encephalitis virus. *J Virol* 70, 8142-8147.
- Stiasny, K., and Heinz, F.X. (2006). Flavivirus membrane fusion. *J Gen Virol* 87, 2755-2766.
- Stocks, C.E., and Lobigs, M. (1998). Signal peptidase cleavage at the flavivirus C-prM junction: dependence on the viral NS2B-3 protease for efficient processing requires determinants in C, the signal peptide, and prM. *J Virol* 72, 2141-2149.
- Storoni, L.C., McCoy, A.J., and Read, R.J. (2004). Likelihood-enhanced fast rotation functions. *Acta Crystallogr D Biol Crystallogr* 60, 432-438.
- Story, R.M., Li, H., and Abelson, J.N. (2001). Crystal structure of a DEAD box protein from the hyperthermophile Methanococcus jannaschii. *Proc Natl Acad Sci U S A* 98, 1465-1470.
- Story, R.M., and Steitz, T.A. (1992). Structure of the recA protein-ADP complex. *Nature* 355, 374-376.
- Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* 41, 207-234.
- Suzich, J.A., Tamura, J.K., Palmer-Hill, F., Warrenner, P., Grakoui, A., Rice, C.M., Feinstone, S.M., and Collett, M.S. (1993). Hepatitis C virus NS3 protein

References

- polynucleotide-stimulated nucleoside triphosphatase and comparison with the related pestivirus and flavivirus enzymes. *J Virol* 67, 6152-6158.
- Svergun, D.I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Cryst.* 25, 495-503.
- Svergun, D.I. (1994). Solution scattering from biopolymers: advanced contrast-variation data analysis. *Acta Cryst.* A50, 391-402.
- Svergun, D.I. (1997). Restoring Three-Dimensional Structure of Biopolymers from Solution Scattering. *J. Appl. Cryst.* 30, 792-797
- Svergun, D.I., Petoukhov, M.V., and Koch, M.J. (2001). Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* 80, 2946-2953.
- Tai, C.L., Chi, W.K., Chen, D.S., and Hwang, L.H. (1996). The helicase activity associated with hepatitis C virus nonstructural protein 3 (NS3). *J Virol* 70, 8477-8484.
- Tai, C.L., Pan, W.C., Liaw, S.H., Yang, U.C., Hwang, L.H., and Chen, D.S. (2001). Structure-based mutational analysis of the hepatitis C virus NS3 helicase. *J Virol* 75, 8289-8297.
- Tan, B.H., Fu, J., Sugrue, R.J., Yap, E.H., Chan, Y.C., and Tan, Y.H. (1996). Recombinant dengue type 1 virus NS5 protein expressed in *Escherichia coli* exhibits RNA-dependent RNA polymerase activity. *Virology* 216, 317-325.
- Tanner, N.K., Cordin, O., Banroques, J., Doere, M., and Linder, P. (2003). The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis. *Mol Cell* 11, 127-138.
- Tanner, N.K., and Linder, P. (2001). DExD/H box RNA helicases: from generic motors to specific dissociation functions. *Mol Cell* 8, 251-262.
- Teo, K.F., and Wright, P.J. (1997). Internal proteolysis of the NS3 protein specified by dengue virus 2. *J Gen Virol* 78 (Pt 2), 337-341.
- Theis, K., Chen, P.J., Skorvaga, M., Van Houten, B., and Kisker, C. (1999). Crystal structure of UvrB, a DNA helicase adapted for nucleotide excision repair. *EMBO J* 18, 6899-6907.
- Turner, C., Witwer, C., Hofacker, I.L., and Stadler, P.F. (2004). Conserved RNA secondary structures in Flaviviridae genomes. *J Gen Virol* 85, 1113-1124.
- Tilgner, M., and Shi, P.Y. (2004). Structure and function of the 3' terminal six nucleotides of the west nile virus genome in viral replication. *J Virol* 78, 8159-8171.
- Tuteja, N., and Tuteja, R. (2004). Unraveling DNA helicases. Motif, structure, mechanism and function. *Eur J Biochem* 271, 1849-1863.
- Uchil, P.D., and Satchidanandam, V. (2003a). Architecture of the flaviviral replication complex. Protease, nuclease, and detergents reveal encasement within double-layered membrane compartments. *J Biol Chem* 278, 24388-24398.

References

- Uchil, P.D., and Satchidanandam, V. (2003b). Characterization of RNA synthesis, replication mechanism, and in vitro RNA-dependent RNA polymerase activity of Japanese encephalitis virus. *Virology* 307, 358-371.
- Umareddy, I., Chao, A., Sampath, A., Gu, F., and Vasudevan, S.G. (2006). Dengue virus NS4B interacts with NS3 and dissociates it from single-stranded RNA. *J Gen Virol* 87, 2605-2614.
- Valle, R.P., and Falgout, B. (1998). Mutagenesis of the NS3 protease of dengue virus type 2. *J Virol* 72, 624-632.
- Varnavski, A.N., and Khromykh, A.A. (1999). Noncytopathic flavivirus replicon RNA-based system for expression and delivery of heterologous genes. *Virology* 255, 366-375.
- Vasilakis, N., and Weaver, S.C. (2008). The history and evolution of human dengue emergence. *Adv Virus Res* 72, 1-76.
- Velankar, S.S., Soultanas, P., Dillingham, M.S., Subramanya, H.S., and Wigley, D.B. (1999). Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell* 97, 75-84.
- Villordo, S.M., and Gamarnik, A.V. (2008). Genome cyclization as strategy for flavivirus RNA replication. *Virus Res.*
- von Moeller, H., Basquin, C., and Conti, E. (2009). The mRNA export protein DBP5 binds RNA and the cytoplasmic nucleoporin NUP214 in a mutually exclusive manner. *Nat Struct Mol Biol.*
- Walker, J.E., Saraste, M., Runswick, M.J., and Gay, N.J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J* 1, 945-951.
- Wallis, T.P., Huang, C.Y., Nimkar, S.B., Young, P.R., and Gorman, J.J. (2004). Determination of the disulfide bond arrangement of dengue virus NS1 protein. *J Biol Chem* 279, 20729-20741.
- Wang, C.C., Huang, Z.S., Chiang, P.L., Chen, C.T., and Wu, H.N. (2009). Analysis of the nucleoside triphosphatase, RNA triphosphatase, and unwinding activities of the helicase domain of dengue virus NS3 protein. *FEBS Lett.*
- Wang, E., Ni, H., Xu, R., Barrett, A.D., Watowich, S.J., Gubler, D.J., and Weaver, S.C. (2000). Evolutionary relationships of endemic/epidemic and sylvatic dengue viruses. *J Virol* 74, 3227-3234.
- Wardell, A.D., Errington, W., Ciaramella, G., Merson, J., and McGarvey, M.J. (1999). Characterization and mutational analysis of the helicase and NTPase activities of hepatitis C virus full-length NS3 protein. *J Gen Virol* 80 (Pt 3), 701-709.
- Wengler, G. (1981). Terminal sequences of the genome and replicative-form RNA of the flavivirus West Nile virus: absence of poly(A) and possible role in RNA replication. *Virology* 113, 544-555.

References

- Wengler, G., Czaya, G., Farber, P.M., and Hegemann, J.H. (1991). In vitro synthesis of West Nile virus proteins indicates that the amino-terminal segment of the NS3 protein contains the active centre of the protease which cleaves the viral polyprotein after multiple basic amino acids. *J Gen Virol* 72 (Pt 4), 851-858.
- Westaway, E.G., Mackenzie, J.M., and Khromykh, A.A. (2003). Kunjin RNA replication and applications of Kunjin replicons. *Adv Virus Res* 59, 99-140.
- WHO (2009). Fact sheet: Dengue and dengue haemorrhagic fever. In WHO fact sheet :Dengue and dengue haemorrhagic fever.
- Winkler, G., Maxwell, S.E., Ruebmler, C., and Stollar, V. (1989). Newly synthesized dengue-2 virus nonstructural protein NS1 is a soluble protein but becomes partially hydrophobic and membrane-associated after dimerization. *Virology* 171, 302-305.
- Winkler, G., Randolph, V.B., Cleaves, G.R., Ryan, T.E., and Stollar, V. (1988). Evidence that the mature form of the flavivirus nonstructural protein NS1 is a dimer. *Virology* 162, 187-196.
- Wu, C.F., Wang, S.H., Sun, C.M., Hu, S.T., and Syu, W.J. (2003). Activation of dengue protease autocleavage at the NS2B-NS3 junction by recombinant NS3 and GST-NS2B fusion proteins. *J Virol Methods* 114, 45-54.
- Wu, J., Bera, A.K., Kuhn, R.J., and Smith, J.L. (2005). Structure of the Flavivirus helicase: implications for catalytic activity, protein interactions, and proteolytic processing. *J Virol* 79, 10268-10277.
- Xu (2007). X-ray crystallographic studies of the Dengue virus helicase/nucleoside triphosphatase catalytic domain. PhD Thesis, School of Biological Sciences, Nanyang Technological University, Singapore.
- Xu, T., Sampath, A., Chao, A., Wen, D., Nanao, M., Chene, P., Vasudevan, S.G., and Lescar, J. (2005). Structure of the Dengue virus helicase/nucleoside triphosphatase catalytic domain at a resolution of 2.4 Å. *J Virol* 79, 10278-10288.
- Xu, T., Sampath, A., Chao, A., Wen, D., Nanao, M., Luo, D., Chene, P., Vasudevan, S.G., and Lescar, J. (2006). Towards the design of flavivirus helicase/NTPase inhibitors: crystallographic and mutagenesis studies of the dengue virus NS3 helicase catalytic domain. *Novartis Found Symp* 277, 87-97; discussion 97-101, 251-103.
- Yamashita, T., Unno, H., Mori, Y., Tani, H., Moriishi, K., Takamizawa, A., Agoh, M., Tsukihara, T., and Matsuura, Y. (2008). Crystal structure of the catalytic domain of Japanese encephalitis virus NS3 helicase/nucleoside triphosphatase at a resolution of 1.8 Å. *Virology* 373, 426-436.
- Yan, Y., Li, Y., Munshi, S., Sardana, V., Cole, J.L., Sardana, M., Steinkuehler, C., Tomei, L., De Francesco, R., Kuo, L.C., and Chen, Z. (1998). Complex of NS3 protease and NS4A peptide of BK strain hepatitis C virus: a 2.2 Å resolution structure in a hexagonal crystal form. *Protein Sci* 7, 837-847.
- Yang, Q., Del Campo, M., Lambowitz, A.M., and Jankowsky, E. (2007). DEAD-box proteins unwind duplexes by local strand separation. *Mol Cell* 28, 253-263.

References

- Yao, N., Hesson, T., Cable, M., Hong, Z., Kwong, A.D., Le, H.V., and Weber, P.C. (1997). Structure of the hepatitis C virus RNA helicase domain. *Nat Struct Biol* 4, 463-467.
- Yao, N., Reichert, P., Taremi, S.S., Prosise, W.W., and Weber, P.C. (1999). Molecular views of viral polyprotein processing revealed by the crystal structure of the hepatitis C virus bifunctional protease-helicase. *Structure* 7, 1353-1363.
- Yap, T.L., Xu, T., Chen, Y.L., Malet, H., Egloff, M.P., Canard, B., Vasudevan, S.G., and Lescar, J. (2007). Crystal structure of the dengue virus RNA-dependent RNA polymerase catalytic domain at 1.85-angstrom resolution. *J Virol* 81, 4753-4765.
- Yon, C., Teramoto, T., Mueller, N., Phelan, J., Ganesh, V.K., Murthy, K.H., and Padmanabhan, R. (2005). Modulation of the nucleoside triphosphatase/RNA helicase and 5'-RNA triphosphatase activities of Dengue virus type 2 nonstructural protein 3 (NS3) by interaction with NS5, the RNA-dependent RNA polymerase. *J Biol Chem* 280, 27412-27419.
- You, S., Falgout, B., Markoff, L., and Padmanabhan, R. (2001). In vitro RNA synthesis from exogenous dengue viral RNA templates requires long range interactions between 5'- and 3'-terminal regions that influence RNA structure. *J Biol Chem* 276, 15581-15591.
- You, S., and Padmanabhan, R. (1999). A novel in vitro replication system for Dengue virus. Initiation of RNA synthesis at the 3'-end of exogenous viral RNA templates requires 5'- and 3'-terminal complementary sequence motifs of the viral RNA. *J Biol Chem* 274, 33714-33722.
- Yu, I.M., Zhang, W., Holdaway, H.A., Li, L., Kostyuchenko, V.A., Chipman, P.R., Kuhn, R.J., Rossmann, M.G., and Chen, J. (2008a). Structure of the immature dengue virus at low pH primes proteolytic maturation. *Science* 319, 1834-1837.
- Yu, L., Nomaguchi, M., Padmanabhan, R., and Markoff, L. (2008b). Specific requirements for elements of the 5' and 3' terminal regions in flavivirus RNA synthesis and viral replication. *Virology* 374, 170-185.
- Yusof, R., Clum, S., Wetzel, M., Murthy, H.M., and Padmanabhan, R. (2000). Purified NS2B/NS3 serine protease of dengue virus type 2 exhibits cofactor NS2B dependence for cleavage of substrates with dibasic amino acids in vitro. *J Biol Chem* 275, 9963-9969.
- Zeng, L., Falgout, B., and Markoff, L. (1998). Identification of specific nucleotide sequences within the conserved 3'-SL in the dengue type 2 virus genome required for replication. *J Virol* 72, 7510-7522.
- Zhang, B., Dong, H., Stein, D.A., Iversen, P.L., and Shi, P.Y. (2008). West Nile virus genome cyclization and RNA replication require two pairs of long-distance RNA interactions. *Virology* 373, 1-13.

References

- Zhang, L., Mohan, P.M., and Padmanabhan, R. (1992). Processing and localization of Dengue virus type 2 polyprotein precursor NS3-NS4A-NS4B-NS5. *J Virol* 66, 7549-7554.
- Zhang, L., and Padmanabhan, R. (1993). Role of protein conformation in the processing of dengue virus type 2 nonstructural polyprotein precursor. *Gene* 129, 197-205.
- Zhang, W., Chipman, P.R., Corver, J., Johnson, P.R., Zhang, Y., Mukhopadhyay, S., Baker, T.S., Strauss, J.H., Rossmann, M.G., and Kuhn, R.J. (2003). Visualization of membrane protein domains by cryo-electron microscopy of dengue virus. *Nat Struct Biol* 10, 907-912.
- Zheng, W., Liao, J.C., Brooks, B.R., and Doniach, S. (2007). Toward the mechanism of dynamical couplings and translocation in hepatitis C virus NS3 helicase using elastic network model. *Proteins* 67, 886-896.
- Zhou, Y., Ray, D., Zhao, Y., Dong, H., Ren, S., Li, Z., Guo, Y., Bernard, K.A., Shi, P.Y., and Li, H. (2007). Structure and function of flavivirus NS5 methyltransferase. *J Virol* 81, 3891-3903.

Appendix

A: Solutions

buffer A (20 mM Na₃PO₄, pH 7.4, 0.5 M NaCl, 40 mM imidazole)

buffer B (20 mM Na₃PO₄, pH 7.4, 0.2 M NaCl)

modified PBS buffer (200 mM NaCl, 10 mM Phosphate, 2.7 mM KCl, and a pH of 7.4)

buffer C (20 mM Tris-Cl, pH 8.0, 0.05 M NaCl)

buffer D (20 mM Tris-HCl, pH 7.4, 0.2 M NaCl, 1 mM dithiothreitol, 5% Glycerol)

buffer D' (20 mM Tris-HCl, pH 7.4, 0.15 M NaCl, 5 mM dithiothreitol, 5% Glycerol)

buffer D'' (20 mM Tris-HCl, pH 7.4, 50mM NaCl, 5 mM β-mercaptoethanol, 5% Glycerol)

B: Proteolysis of scNS2B₁₈NS3.

Band1

Cycle number	Main Residue/s	Minor residue/s	Comments
1	G		RESIDUE1(NS2B)- GSAMAD MW=72.3KDA
2	S		
3	A		
4	M		
5	A		
6	D		

Band2

Cycle number	Main Residue/s	Minor residue/s	Comments
1		D, G, E, T	RESIDUE13(NS3)-13- TQKATL MW=68.3KDA
2	Q		
3			
4	A		
5	T		
6	L		

Band3

Cycle number	Main Residue/s	Minor residue/s	Comments
1	I, E	A	RESIDUE171(NS3)-- IGEPDYE MW=51.2KDA
2	G, R	P, V	
3	E, I		
4	P, G		
5	E, D		
6	Y		
7	D, E		

Band 4

Cycle number	Main Residue/s	Minor residue/s	Comments
1	S	G	RESIDUE249(NS3)-- SDHTG MW=42.2KDA
2	D		
3	H	A	
4	T		
5			
6	R	V, E, K	

C: Sequence Alignment of the flavivirus NS3 protein

CLUSTAL 2.0.11

DEN4	SG-ALMDVPSPAATQKAT-LSEGVRIMCRG-LFGFTQGVGVIHMEGVFHTMMHVTGSGVICHETGRLEPSWADVNDMI SYGGGKLGDIWDYEDQVLAIEPGINPKHVQTHPGLFK 117
DEN1	SG-VLMDTSPPEVEFAV-LDDGIYILCRG-LLGHSQVGVGFQGVFHTMMHVTGAVLMHQKRLPEPSWASVHTDLISYGGKLFQSGWNTGEEVQVIAVEPGINPNVQTPATFTF 117
DEN2	AG-VLMDVSPPPVGFAB-LEDGAYIKQIC-ILGYSQIGAGVFEGFHTMMHVTGAVLMHFGKRIEPSWADVHTDLISYGGKLFLEGEHGEEOVQVLAIEPGINPRAVQTHPGLFK 117
DEN3	SG-VLMDVSPPEVAFAB-LEEGVYIKQGS-IFGFTQGVGQVGFEGFHTMMHVTGAVLTNGKRLPEPSWASVETDLISYGGGKLSAQWQGEVQVIAVEPGINPNFQTMPGIFQ 117
KUN	GG-VLMDTSPREYERGD-TTIGVYIMTRG-LLGSYQAGVWVEGVFHTLMHTTGAAIMSGEGKLDPEYWSVIEDKLCYGGPWILQHFWNGCDEVQMVVVEPGINVNVTTPGVFK 117
MVEV	GG-VEMDTPSPVYPKGD-TTPGVYIMARG-ILGRYQAGVWMEGVFHTLMHTTGAAIMSGEGRLTFYWSNVEDKVTYGGPWILDQFWNGVDDVQMVVVEPGIPAINVQTHPGLFK 117
SLEV	GG-ALMDVSPFVYPKCE-TTPGIYIMTRG-ILGTFQAGVWMEGVFHTMMHVTGAVLRNDEGKLDPEYASDVNDLISYGGPWILSATWDGTTEEQVQVIAVPGIPAINVQTHPGLFK 117
JEV	GG-VEMDTPSPKPLGD-TTIGVYIMARG-ILGTQAGVWMEYENVEFHTLMHTTGAAIMSGEGKLTPEYWSVEDKISYGGPWIFDKWNGTDDVQVIVVEPGIPAVNIQTHPGVFK 117
YFV	SGDLMDIPTPHIEECHEHLEGGIYIFQST-FLGASQGVGVAQGGVFHTMMHVTGAFIVRNGFTLIPSWASVEDLVAYGGSWLEGEWDSGEEVQLIAAPPGINVNVTTPESLFK 119
TBEV	SDLVPSGGGRERGRPEVIGVYRIFSPGLFWQGNQGVGYGKGVLTMMHVTGAAALISIDDAVAIFYWADVEDVVCVGASLEEWKSG-ETVQVHAEPFPGRAHEVHQCPQELI 119

DEN4	TLTG-EIGAVTLDFPPTSGSPIINFKGVGLGVNGVWTKSGDYVSATIQ--AEI-LGEPDYENDE-DIFRFRRLTIMDLHPGAGTILILPSTIVREALKRRLTLILAPTRVVAEME 232
DEN1	TPEG-EVGALALDFPPTSGSPIVNRBEGVGLGVNGVWTTSGTVSAIAQ--AFASQEGPPEIED-EVFRFRNLTIMDLHPGSGTTRKYLPAIVREAIKRLTLVLAPTRVVAASEMA 233
DEN2	TNTG-TIGAVSLDFPPTSGSPIINFKGVGLGVNGVWTTSGTVSAIAQ--TELSIED-NPEIED-DIFRFRRLTIMDLHPGAGTTRKYLPAIVREAIKRLTLILAPTRVVAEME 232
DEN3	TTTG-EIGALALDFPPTSGSPIINBKVGLGVNGVWTTNGTVSIAQ--TNAPDGTPELEE-EMFRFRNLTIMDLHPGSGTTRKYLPAIVREAIKRLTLILAPTRVVAASEMA 233
KUN	TPEG-EIGAVTLDFPPTSGSPIVNRKNGDVLGLGVNGVIMPNGSYISAIVQ--GERMDEPVPAPEP-EMLAFRCITVLDLHPGAGTTRKILPQIIFDAIQRRLTAVLAPTRVVAEMA 233
MVEV	TAHG-EIGAVSLDYPITGSGSPIVNSNGEIIIGLVNGVILGNAYVSAIVQ--GERVEEPVPEAINP-EMLFRQLTVLHLHPGAGTTRKILPQIIFDAIQRRLTAVLAPTRVVAEMA 233
SLEV	TPLG-TIGAVTLDFPPTSGSPIINFKGEIIGLVNGVLIQGEYVSGIIC--GERTEEPIDAYNE-EMLFRRLTVLELHPGAGTTRKVLQIIFDQIQRLTAVLAPTRVVAACEIA 233
JEV	TPFG-EIGAVSLDYPKGTSGSPIILDSNGDIIIGLVNGVELGDGTVSAIVQ--GDACEEPVDATP-SMLFRQMTVLDLHPGSGTTRKILPQIIFDAIQRRLTAVLAPTRVVAEMA 233
YFV	VKNGEIGAVALDYPSTSGSPIVNRNGEVTGLGVNGVILGDSFVSAISQ--TEVIEEGFEELCEIPTMLFRQMTVLDFHPGAGTTRKFLPQILAEARRRLTLVLAPTRVWLESEM 237
TBEV	LDTGRKLGALPIDLVWGTSGSPIINAQGVVWGLYNGGLKT-NETVYSSIAQGEAEISRPNLPCAVVGTGNTSTSGCIIVLDMHPGSGTTHKVLPELLRQCIDRLTLVLAPTRVVLHEME 238

DEN4	EALRGLPIRYQTTPAVRSEHTGR EIVDLMCHATFTTRELSSTRVFNYNLIVMDEAHFTDPASVAAAGYISTRVEM-EAAAIEFMTATPPGSDPPGQSNSPIEDIEREIPERSWNTIGFDWIT 352
DEN1	EALRGLPIRYQTTPAVRSEHTGR EIVDLMCHATFTTRELSSTRVFNYNLIMDEAHFTDPASIAAAGYISTRVEM-EAAAIEFMTATPPGSEMEAFPPQSNNAVINQDEERDIPERSWNTIGFDWIT 353
DEN2	EALRGLPIRYQTTPAIRAEHTGR EIVDLMCHATFTTRELSSTRVFNYNLIMDEAHFTDPASIAAAGYISTRVEM-EAAAIEFMTATPPGSDPPGQSNAPIMDEERDIPERSWNTIGFDWIT 352
DEN3	EALRGLPIRYQTTPATATSEHTGR EIVDLMCHATFTTRELSSTRVFNYNLIMDEAHFTDPASIAAAGYISTRVEM-EAAAIEFMTATPPGSDPPGQSNAPIMDEERDIPERSWNTIGFDWIT 353
KUN	EALRGLPIRYQTSAVAREHNGEIVDMCHATLTHRLMSPLRVFNYNLIVMDEAHFTDPASIAAAGYISTREL-EAAAIEFMTATPPGSDPPGQSNAPISDLQTEIPDRWNSGYEWIT 353
MVEV	EALRGLPVAYLTPAVQREHSNEIVDMCHATLTHRLMSPLRVFNYNLIVMDEAHFTDPASIAAAGYIATRVEM-EAAAIEFMTATPPGSDPPGQSNPVHDVSSEIPDRWSSGPEWIT 353
SLEV	EALRGLPIRYLTPAVREHNGEIVDMCHATLTQLTPTFRVNYQVIIMDEAHFTDPASIAAAGYISTREL-EAAAIEFMTATPPGSDPPGQSNPIIDVEAQVDFKAWSTGIEWIT 353
JEV	EALRGLPVAYTSAVQREHNGEIVDMCHATLTHRLMSPLRVFNYNLIVMDEAHFTDPASIAAAGYIATRVEM-EAAAIEFMTATPPGSDPPGQSNPIIDVEAQVDFKAWSTGIEWIT 353
YFV	EAFHGLDVIPTQAFSAHSGSEVIVDMCHATLTLMLEBTRVFNYNLIVMDEAHFTDPASIAAAGYIATRVEM-EAAAIEFMTATPPGSDPPGQSNPIIDVEAQVDFKAWSTGIEWIT 357
TBEV	EALNGEKVAFHSPAVSDQAGGAIVDMCHATYVNRRLPQGRWEVAINDEAHWTDPASIAAKGHLTLTIAENFCALVLMZTATPPGSEPPFESNGALITSEERKQIPDGENEDGFDWIT 358

DEN4	RPRWLDARVYADPNALVDFEFASGRK	618
DEN1	RPRWLDARTYSDPMALVEFEFAAAGR	619
DEN2	RPRWLDARTYSDPLALVEFEFAAAGR	618
DEN3	RPRWLDARTYSDPLALVEFEFAAAGR	619
DEN KUN	RPRWLDARTYSDHCLALVFKDFASGR	619
WVEV	RPRWLDARTYSDHGLLVFKDFFAAGR	619
SLVEV	RPRWLDARTYSDCCVLCALVFEFAAGR	618
JEV	RPRWLDARTYADHCLLVFKDFFAAGR	619
YFV	RPRWCDENVSSDQSLVSEFKFAEGR	623
BEV	RPRWCDARMEKEGRLVEFVYASGR	621

D: Crystallographic contacts in common between scNS2B₁₈NS3**Conformation I and II.**

Source atoms	Target atoms	Distance (Å)	Symmetry operation
Lys 186A NZ	Asn 576A ND2	2.99 (2.93)	1: -X, Y+1/2, -Z
Leu 188A N	Asn 576A O	2.87 (3.01)	1: -X, Y+1/2, -Z
Lys 248A N	Glu 336A O	3.10 (2.73)	2: X, Y, Z
Ser 249A O	Glu 336A N	2.75 (2.94)	2: X, Y, Z
Asp 250A OD2	Glu 333A OE2	2.96 (2.70)	2: X, Y, Z
His 251A O	Asp 334A N	2.97 (2.82)	2: X, Y, Z
Arg 254A NH1	Asp 469A OD2	2.76 (3.16)	2: X, Y, Z
	(Glu 468A OE2)		
Ser 272A OG	Glu 336A OE1	2.51 (2.70)	2: X, Y, Z
Glu 305A OE2	Asn 576A OD1	3.02 (3.06)	1: -X, Y+1/2, -Z
Glu 333A OE2	Asp 250A OD2	2.96 (2.70)	2: X, Y, Z
Asp 334A N	His 251A O	2.97 (2.82)	2: X, Y, Z
Glu 336A N	Ser 249A O	2.75 (2.94)	2: X, Y, Z
Glu 336A OE1	Ser 249A OG	2.94 (2.85)	2: X, Y, Z
Glu 336A OE1	Ser 272A OG	2.51 (2.70)	2: X, Y, Z
Glu 336A O	Lys 248A N	3.10 (2.73)	2: X, Y, Z
Glu 338A OE1	Arg 539A NH2	3.17 (3.16)	2: X, Y, Z
Asp 469A OD2	Arg 254A NH1	2.76 (3.16)	2: X, Y, Z
(Glu 468A OE2) ¹			
Arg 513A NH2	Glu 575A OE1	3.11 (2.79)	1: -X, Y+1/2, -Z
Arg 539A NH1	Glu 338A OE1	3.06 (3.16)	2: X, Y, Z
Glu 575A OE1	Arg 513A NH2	3.11 (2.79)	1: -X, Y+1/2, -Z
Asn 576A OD1	Glu 305A OE2	3.02 (3.06)	1: -X, Y+1/2, -Z
Asn 576A ND2	Lys 186A NZ	2.99 (2.93)	1: -X, Y+1/2, -Z
Asn 576A O	Leu 188A N	2.87 (3.01)	1: -X, Y+1/2, -Z

¹ The atoms and distances from Conformation II are in parentheses.

E: CV_ Dahai Luo

Curriculum Vitae

Date of preparation: Oct 1st, 2009

Dahai Luo

E-mail: lu0001ai@ntu.edu.sg, dahai_luo@hotmail.com.

Office phone: +65-63162858

Address: 03S-51, School of Biological Sciences, Nanyang Technological University,
60 Nanyang Drive, Singapore 637551.

Education

Aug 2006 -
present	Graduate student.
Supervisor: Dr Julien Lescar
Major: Protein Science and Macromolecular X-ray Crystallography
Research topic: Structural and functional studies of the viral
replication enzymes from flaviviruses (focusing on: NonStructural
protein 3 – NS3)
School of Biological Sciences, Nanyang Technological University,
Singapore.
Jul 2002 - Jun
2006	B.Sc. (Hons)
School of Biological Sciences, Nanyang Technological University,
Singapore.
Jan - Jun 2002	Pre-university Bridging Course.
National Institute of Education (NIE), Nanyang Technological
University, Singapore.

Publications

- 1: **Luo D**, Xu T, Watson R, Scherer-Becker D, Sampath A, Jahnke W, Yeong SS, Wang CH, Lim SP, Strongin A, Vasudevan SG, Lescar J.
Insights into RNA unwinding and ATP hydrolysis by the flavivirus NS3 protein.
EMBO J. 2008 Dec 3; 27(23):3209-19.
- 2: Lescar J, **Luo D**, Xu T, Sampath A, Lim SP, Canard B, Vasudevan SG.
Towards the design of antiviral inhibitors against flaviviruses:
The case for the multifunctional NS3 protein from Dengue virus as a target.
Antiviral Res. 2008 Nov; 80(2):94-101.
- 3: **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 Jan; 82(1):173-83.

4: Xu T, Sampath A, Chao A, Wen D, Nanao M, **Luo D**, Chene P, Vasudevan SG, Lescar J.

Towards the design of flavivirus helicase/NTPase inhibitors: crystallographic and mutagenesis studies of the dengue virus NS3 helicase catalytic domain.

Novartis Found Symp. 2006; 277:87-97; discussion 97-101, 251-3.

5: Sampath A, Xu T, Chao A, **Luo D**, Lescar J, Vasudevan SG.

Structure-based mutational analysis of the NS3 helicase from dengue virus.

J Virol. 2006 Jul; 80(13):6686-90.

Scientific Experiences & Awards

08-11 Dec 2009	Talk: The NS3 Protein from Dengue Virus: Structure and Function
Emerging Infectious Diseases Conference 2009, Singapore.
27 Jun -02 Jul 2009	Poster:
EMBO conference "Helicases and NTP-Driven Nucleic Acid Motors" Les Diablerets, Switzerland.
23-31 Aug 2008	Talk: Crystal structure of the NS3 protease-helicase from Dengue virus
Microsymposium 79 "Motion in macromolecular machines"
XXI Congress and General Assembly of the International Union of Crystallography (IUCr). Osaka, Japan.
21-24 Sep 2007	Student Travel Scholarship
Poster: Structural and Functional studies of the NS3 protease-helicase from Dengue virus
3rd Asian Regional Dengue Research Network Meeting, Taipei, China.
Aug 2006- Aug 2007	Research attachment.
Project title: Crystal structure of the NS3 protease-helicase from dengue virus
Supervisors: Dr Subhash G Vasudevan and Dr Julien Lescar
Dengue Department, Novartis Institute for Tropical Diseases, Biopolis, Singapore.
28-30 Sep 2005	Student Travel Scholarship
2nd Asian Regional Dengue Research Network Meeting, Singapore.
May-Jul 2005	Research attachment.
Project title: Structure-based mutational analysis of the NS3 helicase from dengue virus
Supervisors: Dr Aruna Sampath and Dr Subhash G Vasudevan
Dengue Department, Novartis Institute for Tropical Diseases,

Biopolis, Singapore.

Nov-Dec 2004

Research attachment.
Project title: Structural and functional studies of the SARS
3CL protease
Supervisors: Dr Julien Lescar
School of Biological Sciences, Nanyang Technological
University, Singapore.

Techniques

Molecular biology	Cloning (PCR) Site-directed mutagenesis Purification of plasmid DNA DNA sequencing cDNA synthesis
Protein preparation and characterization	Recombinant protein expression using bacterial system Purification of tagged and untagged proteins using size exclusion, affinity, hydrophobic, ion exchange chromatography (AKTA systems) SDS-PAGE, Native PAGE, Western-blotting Circular Dichroism Dynamic light scattering Kinetic assays using spectrometric and fluorometric methods Thermofluor
X-ray crystallography	Protein crystallisation screening Co-crystallisation of protein with substrates, inhibitors or novel chemicals Optimization of crystallization conditions Cryo-crystallography Usage of in-house X-ray sources and synchrotron beamlines Solution of protein structure by molecular replacement Refinement and rebuilding of protein structural models
X-ray crystallography related softwares	General: CCP4 suite Data processing: Mosflm, HKL package Solution of structure and refinement: Refmac, CNS Visualization and structure rebuilding: COOT Preparation of figures: Pymol, VMD
General software/OS	Microsoft Office (Word, Excel, PowerPoint) Adobe Photoshop Unix/Linux

Personal

Character: Independent; With good teamwork ability; Initiative; Willing to learn and to take challenges.

Hobbies: Reading, Sports, Travel and Photography.

Reference

References will be provided upon request.

Referees:

Julien Lescar, Ph. D.

Associate Professor

School of Biological Sciences

Nanyang Technological University

60, Nanyang Drive, Singapore 637551

E-mail: julien@ntu.edu.sg

Phone (office): +65-63162859

Fax: +65-67913856

<http://www.ntu.edu.sg/sbs/staff/julien/labsite/Home.htm>

Subhash Vasudevan, Ph.D.

Associate Professor

Program in Emerging Infectious Diseases

Duke-NUS Graduate Medical School

National University of Singapore

E-mail: subhash.vasudevan@duke-nus.edu.sg

Phone (office): +65-65166718

Fax: +65-65348632