NANYANG
TECHNOLOGICAL
UNIVERSITY

**Solution X-ray Scattering Studies on the Dengue Virus Non-
structural Protein 5 and -3 as well as the Capsid Protein**

SAW WUAN GEOK
SCHOOL OF BIOLOGICAL SCIENCES
2017

**Solution X-ray Scattering Studies on the Dengue Virus
Non-structural Protein 5 and -3 as well as the Capsid
Protein**

SAW WUAN GEOK

SCHOOL OF BIOLOGICAL SCIENCES

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

2017

Acknowledgement

I would like to express my deepest gratitude to my advisor Professor Dr. Gerhard Grüber for giving me the opportunity to work on this interesting project, and introducing me to the field of small-angle X-ray scattering. I thank him for all his supports, encouragement and inspiring advices throughout this research work. I am deeply grateful to Dr. Ardina Grüber, Dr. Cornelia Hunke and Dr. Phat Vinh Dip for all their guidance, discussion and suggestions in molecular cloning and protein purification, as well as the fun and joy that they have been provided throughout the years. I am grateful and happy to work with Dr. Malathy Sony Subramanian Manimekalai for all her help in performing small-angle X-ray scattering experiments, and providing a lot of helpful advices and suggestions for the SAXS data analysis and comments to this research work and thesis writing. I thank Ankita Pan for her great effort in helping me with the molecular cloning and protein purification, as well as my thesis correction. I would like to thank Mun Fong Sek for all the discussions that we have, and helping me in the correction of this thesis. I would like to thank Dr. Asha Balakrishna, Dr. Hendrik Sielaff, Dr. Yew Kwang Toh, Dr. Giancarlo Tria, Dr. Anil Kumar and Dr. Wilson Nartey for their generous sharing of knowledge and suggestions in different research techniques. I would like to extend my thankful to Dr. Goran Biukovic, Dr. Kamariah Neelagandan, Dr. Sandip Basak, Dr. Shin Joon, Dr. Lavanya Sundararaman, Ms. Claudia Karin Beushausen geb. Schaefer, Dhirendra Singh, Adam Hotra, Arvind Kumar and Bharti Singal for their assistants and supports.

Beside my colleagues, I am extremely thankful to my Thesis Advisory Committee members Prof. Salil Bose and A/Prof. Susana Geifman Shochat for their insightful comments and advice on this thesis. I would like to thank Prof. Subhash Vasudevan, Dr. Moon Yue Feng Tay, Dr. Yongqian Zhao and Asst. Prof. Dahai Luo for their helps in providing the plasmids of NS5FL, NS3 and helicase constructs. I thank Dr. Feng Gu, Dr. Pei Yong Shi and Dr. Siew Pheng Lim (Novartis Institute for Tropical Diseases) for providing Gppp-RNA (GpppAGUUGUU) and cap-0 RNA (m^7 GpppAGUUGUU). I would like to thank A/Prof. Ding Xiang Liu in providing viral cDNA encoding DENV-2 genomic RNA. I am very grateful for the financial support provided by Nanyang Technological University research scholarship. This research was supported by the Ministry of Education (MOE), Singapore (MOE2012-T3-1-008; G. Grüber). I would like to express my sincere appreciation to Dr. Thomas M. Weiss, Dr. Tsutomu Matsui and Dr. Ivan Rajkovic at beamline 4-2 of the Stanford

Synchrotron Radiation Lightsource (SSRL) for their expert help with TR-SAXS data collection. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research and by the National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393) and the National Center for Research Resources (P41RR001209).

Lastly, I would like to express my special appreciation and thanks to my family for all their supports and sacrifices that they have made for me. I thank all my friends who support me all the time, and providing me a lot of fun throughout the years. A special thanks to my lovely “kid” Chui Fann Wong for all the fun and joy that she brought to me, especially during my thesis writing.

Acknowledgements	I
Table of contents	III
Lists of figures	VI
List of table	VIII
Abbreviations	IX
Abstract	XIII
1. Introduction	1
1.1 Flavivirus	3
1.2 Dengue virus (DENV)	3
1.3 Dengue virus life cycle	4
1.4 DENV virion structure	6
1.5 DENV genome replication	8
1.6 RNA cyclization in DENV genome replication	12
1.7 Non-structural protein 5 (NS5)	13
1.7.1 Methyltransferase (MTase)	15
1.7.2 RNA-dependent RNA polymerase (RdRp)	17
1.7.3 10-residue linker	19
1.8 Interaction of NS3 and NS5	20
1.9 Non-structural protein 3 (NS3)	21
1.9.1 Protease	22
1.9.2 Helicase	24
1.10 NS3 and NS5 as a target for DENV vaccine or antiviral drug development	25
1.11 Capsid protein	26
1.12 Small-angle X-ray scattering (SAXS)	28
1.13 Goals of this thesis	30
2. Materials and Methods	33
2.1 Materials	35
2.1.1 Chemicals	35
2.1.2 Molecular biology materials	35
2.1.3 Chromatography	35
2.1.3.1 Affinity chromatography	
2.1.3.2 Size exclusion chromatography	
2.1.3.3 Instruments and accessories	
2.1.3.4 Protein concentration estimation	
2.1.4 Other instruments	36
2.1.5 Computer software	36
2.2 Methods	37
2.2.1 Vector preparation (modified pET-24b)	37
2.2.2 Cloning of genes encoding viral proteins from DENV	38
2.2.2.1 Cloning of genes of NS5 from DENV	38
2.2.2.2 Cloning of genes encoding NS5 linker mutants from DENV-3 and DENV-4	40
2.2.2.3 Cloning of gene encoding DENV-2 capsid protein	43
2.2.2.4 Information of genes encoding DENV-4 NS3 protein and its domains	43
2.2.3 Protein production of recombinant viral proteins from DENV	44
2.2.3.1 Induction test of viral proteins from DENV	44
2.2.3.2 Protein production of viral proteins from DENV	44
2.2.4 Protein purification of viral proteins from DENV	44

2.2.4.1	Protein purification of NS5 constructs from DENV	45
2.2.4.2	Protein purification of NS3 constructs from DENV-4	45
2.2.4.3	Protein purification of capsid protein from DENV-2	46
2.2.5	Small angle X-ray scattering (SAXS) experiments	46
2.2.5.1	SAXS data collection using home source NANOSTAR	47
2.2.5.2	High-resolution SAXS data collection at Stanford Synchrotron Radiation Lightsource (SSRL)	48
2.2.5.3	Time-resolved SAXS (TR-SAXS) data collection at SSRL	49
2.2.6	Small angle X-ray scattering (SAXS) data analysis	50
2.2.6.1	Primary SAXS data processing	50
2.2.6.2	<i>Ab initio</i> modeling using SAXS	52
2.2.6.3	Protein folding and flexibility assessment using SAXS	53
2.2.6.4	Rigid-body modeling using SAXS	54
2.2.6.5	Ensemble modeling using SAXS	55
2.2.7	Protein sequence analysis	56
3.	Results	57
3.1	Sequence analysis of NS5	59
3.1.1	Sequence analysis of NS5 from DENV-1-4	59
3.1.2	Sequence analysis of NS5 from flaviviruses	62
3.2	Characterization of full-length NS5 (NS5FL) from DENV-3 in solution	65
3.2.1	Expression and purification of DENV-3 NS5FL	66
3.2.2	Primary SAXS data analysis of DENV-3 NS5FL	67
3.2.3	Reconstruction of <i>ab initio</i> models of DENV-3 NS5FL	68
3.2.4	Structural comparison between the SAXS envelope and crystallographic structure	70
3.2.5	Domain mobility of DENV-3 NS5FL	71
3.2.6	SAXS study of individual domains of DENV-3 NS5FL	74
3.3	Characterization of NS5FL from DENV-1 to -4 in solution	77
3.3.1	Expression and purification of NS5FL from DENV-1 to -4	77
3.3.2	Primary SAXS data analysis of NS5FL from DENV-1 to -4	78
3.3.3	Reconstruction of <i>ab initio</i> and rigid-body models of NS5FL from DENV-1 to -4	80
3.3.4	Ensemble analysis of NS5FL from DENV-1 to -4	81
3.4	Mutational studies of the NS5FL linker region of DENV-3 and DENV- 4 using SAXS	84
3.4.1	SAXS study of NS5FL linker mutants based on 10-residues linker substitution	84
3.4.2	NS5 linker mutant design using amino acid substitution	87
3.4.3	SAXS study of NS5FL linker mutants based on amino acid substitution	88
3.4.4	Ensemble analysis of NS5FL linker mutants based on amino acid substitution	90
3.5	Characterization of full-length NS5 (NS5FL) with substrates using SAXS	92
3.5.1	Static SAXS studies of DENV-2 NS5FL with substrates	92
3.5.2	Time-dependent SAXS studies of DENV-4 NS5FL with RNA	95
3.6	Characterization of DENV-4 helicase in solution using SAXS	100
3.6.1	Expression and purification of DENV-4 helicase and hel- NS5 ₃₂₀₋₃₄₁	100

3.6.2	Primary SAXS data analysis of DENV-4 helicase and hel-NS5 ₃₂₀₋₃₄₁	102
3.6.3	<i>Ab initio</i> reconstructions of DENV-4 helicase and hel-NS5 ₃₂₀₋₃₄₁	103
3.6.4	Flexibility assessment of NS5 peptide in hel-NS5 ₃₂₀₋₃₄₁	105
3.7	Characterization of DENV-4 NS2B ₁₈ NS3 in solution using SAXS	107
3.7.1	Expression and purification of DENV-4 NS2B ₁₈ NS3	108
3.7.2	Primary SAXS data analysis of DENV-4 NS2B ₁₈ NS3	109
3.7.3	Reconstruction of <i>ab initio</i> models of DENV-4 NS2B ₁₈ NS3	110
3.7.4	Structural comparison between the SAXS envelope and crystallographic structure	111
3.7.5	Domain mobility of DENV-4 NS2B ₁₈ NS3	112
3.8	Characterization of DENV-2 capsid protein in solution using SAXS	115
3.8.1	Expression and purification of DENV-2 capsid protein	115
3.8.2	Primary SAXS data analysis of DENV-2 capsid protein	116
3.8.3	Reconstruction of rigid-body models and flexibility analysis of the capsid protein from DENV-2	118
3.8.4	Salt effect on the DENV-2 capsid protein	121
4.	Discussion	125
4.1	Domain-domain arrangement of DENV NS5	127
4.2	Conformational flexibility of DENV NS5 in solution	129
4.3	Similarity and diversity of the NS5FLs from DENV-1 to -4	130
4.4	Substrate binding of NS5	133
4.5	NS3-NS5 interacting epitope	134
4.6	Structural characterization of NS3 from DENV-4 in solution	135
4.7	Structural characterization of capsid protein from DENV-2 in solution	137
4.8	Overview of SAXS technique	139
5.	Conclusions	143
6.	References	147
7.	Others	165

Introduction

Figure 1.1	DENV virion structure.	7
Figure 1.2	Schematic diagrams of DENV (<i>top</i>) viral genome and (<i>bottom</i>) polyprotein.	8
Figure 1.3	Schematic diagram of hypothetical model of DENV replication complex.	9
Figure 1.4	The proposed mechanism of DENV RNA replication.	10
Figure 1.5	The proposed mechanism of DENV RNA capping mechanism.	11
Figure 1.6	Schematic diagram of DENV RNA cyclization.	12
Figure 1.7	The MTase domain taken from the crystallographic structure of DENV-3 NS5 ₆₋₈₉₅ -RNA co-crystal (PDB ID: 5DTO)	17
Figure 1.8	The crystallographic structure of DENV-3 NS5 ₂₆₅₋₉₀₀ (RdRp; PDB ID: 4C11)	18
Figure 1.9	Crystallographic structures of DENV-4 NS2B ₁₈ NS3 adopted conformation I (PDB ID: 2VBC) and conformation II (PDB ID: 2WHX).	22
Figure 1.10	The protease domain of DENV-3 NS3 in complex with the NS2B ₅₀₋₉₅ and the inhibitor (PDB ID: 3U1I)	23
Figure 1.11	The helicase domain of DENV-3 NS3 in complex with the ssRNA (12-mer) and the AMPPNP (PDB ID: 2UJV).	24

Materials and methods

Figure 2.2.1	Plasmid pET-24b.	37
Figure 2.2.2	BRUKER NANOSTAR SAXS instrument located in School of Biological Sciences, Nanyang Technological University.	47
Figure 2.2.3	SAXS instrument located in Stanford Synchrotron Radiation Lightsource (SSRL).	48
Figure 2.2.4	Stopped-flow device located at SSRL.	50

Results

Figure 3.1.1	Multiple sequence alignments of full-length NS5 from all four serotypes.	61
Figure 3.1.2	Multiple sequence alignments of NS5 from different flaviviruses.	63
Figure 3.2.1	Crystallographic structure of DENV-3 NS5 ₆₋₈₉₅ (PDB ID: 4V0R).	65
Figure 3.2.2	Crystal packing of 4V0R structure of DENV-3 NS5 ₆₋₈₉₅ .	66
Figure 3.2.3	Purification of DENV-3 NS5FL.	67
Figure 3.2.4	Primary SAXS data analysis of DENV-3 NS5FL.	68
Figure 3.2.5	<i>Ab initio</i> modeling of DENV-3 NS5FL.	69
Figure 3.2.6	Rigid body modeling of DENV-3 NS5FL.	70
Figure 3.2.7	Assessment of protein flexibility in DENV-3 NS5FL.	72
Figure 3.2.8	Ensemble formation of DENV-3 NS5FL.	73
Figure 3.2.9	Primary SAXS data analysis of DENV-3 NS5 domains.	75
Figure 3.2.10	Assessment of protein flexibility in DENV-3 NS5 domains.	76
Figure 3.3.1	The SDS gel showed high purity of NS5FL from DENV-1 to DENV-4.	78
Figure 3.3.2	Primary SAXS data analysis of DENV-1-4 NS5FL.	79
Figure 3.3.3	<i>Ab initio</i> reconstruction and rigid body modeling for NS5FL from DENV-1 to DENV-4.	81
Figure 3.3.4	Assessment of protein flexibility in NS5 of DENV-1 to -4.	82
Figure 3.3.5	EOM analysis of NS5FL from DENV-1 to DENV-4.	83
Figure 3.4.1	Protein purification of NS5FL linker mutants.	84
Figure 3.4.2	SAXS scattering patterns of NS5FL linker mutants.	86
Figure 3.4.3	SAXS analysis of NS5FL linker mutants.	86
Figure 3.4.4	Sequence analysis of NS5FL linker mutants.	88
Figure 3.4.5	SAXS analysis of DENV-4 NS5FL linker mutants.	90
Figure 3.4.6	Ensemble analysis of DENV-4 NS5FL linker mutants.	91
Figure 3.5.1	Primary SAXS analysis of DENV-2 NS5FL in the absence or presence of substrates.	93
Figure 3.5.2	SAXS analysis of DENV-2 NS5FL in the absence or presence of substrates.	94

Figure 3.5.3	Static SAXS analysis of DENV-4 NS5FL in the absence or presence of RNA in two different protein concentrations.	96
Figure 3.5.4	Time-dependent SAXS analysis of DENV-4 NS5FL in the absence or presence of RNA at protein concentration of 2.3 mg/ml.	97
Figure 3.5.5	Time-dependent SAXS analysis of DENV-4 NS5FL in the absence or presence of RNA at protein concentration of 4.5 mg/ml.	98
Figure 3.5.6	TR-SAXS analysis of DENV-4 NS5FL in the absence or presence of RNA.	99
Figure 3.6.1	Schematic diagram of DENV-4 NS3 helicase constructs.	100
Figure 3.6.2	Purification of DENV-4 helicase and hel-NS5 ₃₂₀₋₃₄₁ proteins.	101
Figure 3.6.3	Primary SAXS data analysis of DENV-4 helicase and hel-NS5 ₃₂₀₋₃₄₁ .	103
Figure 3.6.4	<i>Ab initio</i> reconstruction of DENV-4 helicase and hel-NS5 ₃₂₀₋₃₄₁ proteins.	104
Figure 3.6.5	Assessment of NS5 peptide flexibility of DENV-4 hel-NS5 ₃₂₀₋₃₄₁ proteins.	105
Figure 3.7.1	The domain-domain orientation of NS3 from DENV, MVES and HCV.	107
Figure 3.7.2	Purification of DENV-4 NS2B ₁₈ NS3.	108
Figure 3.7.3	Primary SAXS data analysis of DENV-4 NS2B ₁₈ NS3.	109
Figure 3.7.4	<i>Ab initio</i> modeling of DENV-4 NS2B ₁₈ NS3.	111
Figure 3.7.5	Comparison of SAXS structure and crystallographic structure.	112
Figure 3.7.6	Assessment of protein flexibility in DENV-4 NS2B ₁₈ NS3.	113
Figure 3.7.7	EOM analysis of DENV-4 NS2B ₁₈ NS3.	113
Figure 3.8.1	Purification of DENV-2 capsid proteins.	116
Figure 3.8.2	Primary SAXS data analysis of DENV-2 capsid protein.	118
Figure 3.8.3	SAXS scattering profile (<i>green circle</i>) of DENV-2 capsid protein, and the calculated scattering profiles of monomeric (<i>red line</i>) or dimeric (<i>blue line</i>) NMR structure 1R6R	119
Figure 3.8.4	Rigid-body modeling of DENV-2 capsid proteins.	119
Figure 3.8.5	Assessment of protein flexibility in DENV-2 capsid protein.	120
Figure 3.8.6	Primary SAXS data analysis of DENV-2 capsid proteins in buffers contained different ionic strength.	122

Discussion

Figure 4.1.1	Crystallographic structures of (<i>left</i>) DENV-3 NS5 ₆₋₈₉₅ (PDB ID: 4V0R) and (<i>right</i>) JEV NS5 (PDB ID: 4K6M).	127
Figure 4.1.2	Possible arrangement of the MTase and the RdRp domains.	128
Figure 4.3.1	Interaction of the NS5 linker residues (<i>black stick</i>) and the MTase (<i>blue stick</i>) or the RdRp (<i>red stick</i>) analyzed from the 4V0R structure.	132
Figure 4.5.1	Schematic diagram of NS3-NS5 interaction.	134
Figure 4.6.1	Comparison of domain-domain orientations of DENV-4 NS2B ₁₈ NS3 in conformation I (PDB ID: 2VBC) and II (PDB ID: 2WHX).	136
Figure 4.7.1	The NMR structure of DENV-2 capsid (residues 20-100) (PDB ID: 1R6R) represented as <i>cartoon</i> (<i>top</i>) and charge distribution surface (<i>bottom</i>).	138

Materials and methods

Table 2.2.1	Construct information of full-length NS5 from DENV-1 to DENV-4	38
Table 2.2.2	Construct information of NS5 domains from DENV-3 and DENV-4	38
Table 2.2.3	Primers used in polymerase chain reaction (PCR) for NS5 individual domains	38
Table 2.2.4	Components of PCR reaction mixture	39
Table 2.2.5	PCR thermocycler program	39
Table 2.2.6	Components of ligation mixture	40
Table 2.2.7	Primers used in PCR for NS5 linker-swapped mutants	40
Table 2.2.8	Primers used in PCR for DENV-4 NS5 linker mutants with amino acid substitution	41
Table 2.2.9	Components of PCR reaction mixture	42
Table 2.2.10	PCR thermocycler program	42
Table 2.2.11	Buffer conditions used in protein purification	45

Results

Table 3.2.1	Structural parameters for individual domains of DENV-3 NS5FL	74
Table 3.3.1	Structural parameters for NS5FL from DENV-1 to -4	79
Table 3.4.1	Structural parameters for NS5FL mutants D3NS5-d4linker and D4NS5-d3linker	85
Table 3.4.2	Amino acid sequences of DENV-4 NS5FL linker mutants	87
Table 3.4.3	Scattering-derived parameters from Guinier region and $P(r)$ function for wild-type NS5 and DENV-4 NS5FL linker mutants	89
Table 3.5.1	Scattering-derived parameters from Guinier region and $P(r)$ functions for DENV-2 NS5FL in the absence or presence of substrates	93
Table 3.6.1	Structural parameters for DENV-4 helicase and hel-NS5 ₃₂₀₋₃₄₁	102
Table 3.7.1	Structural parameters for DENV-4 NS2B ₁₈ NS3	110
Table 3.8.1	Structural parameters for DENV-2 capsid protein	117
Table 3.8.2	Structural parameters for DENV-2 capsid protein in the presence or absence of NaCl	123

Å	Angström
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
B-factor	Temperature factor
CS	Complementary sequence
Da	Dalton
DENV	Dengue virus
DF	Dengue fever
DHF	Dengue haemorrhagic fever
D_{max}	Maximum particle dimension
dNTP	Deoxy-nucleoside triphosphate
dsRNA	Double-stranded ribonucleic acid
DSS	Dengue shock syndrome
DTT	Dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
EOM	Ensemble Optimization Method
ER	Endoplasmic reticulum
GMP	Guanosine monophosphate
gRNA	Genomic ribonucleic acid
GTase	Guanylyltransferase
GTP	Guanosine triphosphate
HCV	Hepatitis C virus
$I(0)$	Forward intensity
IPTG	Isopropyl-β-D-thiogalactoside
JEV	Japanese encephalitis virus
mRNA	Messenger ribonucleic acid
MTase	Methyltransferase
MVES	Murray Valley encephalitis virus
NaCl	Sodium chloride
NC	Nucleocapsid
Ni-NTA	Nickel-nitrilotriacetic acid
NLS	Nuclear localization sequence

NMR	Nuclear magnetic resonance
NS	Non-structural protein
NS5FL	Full-length non-structural protein 5
NSD	Normalized spatial discrepancy
NTPase	Nucleoside triphosphatase
PCR	Polymerase chain reaction
PDB	Protein Data Bank
phosRNA	Phosphorylated ribonucleic acid
$P(r)$	Pair-distance distribution
RC	Replication complex
RdRp	RNA-dependent RNA polymerase
RF	Replicative form
R_g	Radius of gyration
RI	Replicative intermediate
RNA	Ribonucleic acid
RTPase	RNA 5'-triphosphatase
SAH	S-adenosyl-homocysteine
SAM	S-adenosyl-methionine
SAXS	Small-angle X-ray scattering
SDS	Sodium-dodecyl-sulphate
SEC	Size exclusion chromatography
SL	Stem-loop
SPR	Surface plasmon resonance
ssRNA	Single-stranded ribonucleic acid
TM	Transmembrane
TR-SAXS	Time-resolved small-angle X-ray scattering
UAR	Upstream AUG region
UTR	Untranslated region
WNV	West-Nile virus
YFV	Yellow Fever virus

ABSTRACT

Dengue virus (DENV), which is the causative agent of dengue, is a single positive-stranded RNA virus. The structural information of two DENV non-structural proteins, NS5 and NS3, which play a central role in the genome replication, were gathered using small-angle X-ray scattering (SAXS). SAXS revealed that NS5 from all four serotypes adopt an extended and a slightly flexible conformation in solution, with DENV-4 NS5 being slightly more compact and less flexible than the other three serotypes. Three linker residues, S266, T267 and K271 were identified as the key residues responsible for this compactness. Upon adding RNA into the NS5 solution, an increase in size of NS5 was detected using SAXS. Furthermore, the solution data of DENV-4 NS2B₁₈NS3 discloses a favorable domain-domain arrangement, which allows the binding of ADP-Mn²⁺ by the helicase. In addition, the unassigned N-terminal residues of the DENV-2 capsid protein were analyzed using SAXS, and revealed a highly flexible and extended conformation in solution.

1. INTRODUCTION

1.1 Flavivirus

The International Committee on Taxonomy of Viruses (ICTV) classified viruses from the *Flaviviridae* family into four genera, namely *Flavivirus*, *Hepacivirus*, *Pegivirus* and *Pestivirus* (1). The members of this family share some common features, including the morphology of their virions, their genome structures and virus replication. Despite these similarities, they are different in their biological properties and serological cross-reactivity (2). *Flavivirus* is the largest genus in this family, and based on the antigenic properties of the flaviviruses, they have been classified into 53 species, including Dengue virus (DENV), Japanese encephalitis virus (JEV), West Nile virus (WNV), yellow fever virus (YFV), Zika virus (ZIKV) and tick-borne encephalitis virus (TBEV) (1,3,4). Majority of the flaviviruses are arthropod-borne, mainly transmitted through mosquitoes and ticks. They are enveloped viruses with a single-stranded, positive-sense RNA genome that encodes at least ten viral proteins. Their virions contain an envelope and a nucleocapsid. Their envelopes comprise of a host-derived lipid bilayer with multiple copies of envelope (E) and membrane (M) glycosylated proteins. Beneath this envelope, multiple copies of capsid (C) protein encapsidate the viral genome and form the nucleocapsid (NC).

1.2 Dengue virus (DENV)

Dengue virus (DENV) is one of the members of genus *Flavivirus*, and it can be further sub-categorized into four distinct serotype groups based on their antigenic properties, which are designated as DENV-1, -2, -3 and -4. DENV infection is ranked as the most important mosquito-borne viral disease around the world (5). Its primary epidemic vector is *Aedes aegypti*, and human is one of the hosts for this virus. It is predominantly found in tropical and sub-tropical regions, and about 2.5 billion people around the world are at risk of infection. Each year, approximately 500,000 people diagnosed with dengue haemorrhagic fever (DHF) require hospitalization, and approximately 90% of them are children aged less than five years old (6). The infection of DENV can be asymptomatic, in which the virus carrier does not experience any clinical symptoms or develops any clinical manifestations, including dengue fever (DF) or its severe forms, dengue haemorrhagic fever (DHF) and dengue shock syndrome (DSS) (6).

Dengue fever (DF) is the most common among individuals that experience clinical symptoms after DENV infection. It is characterized by symptoms including fever,

headache, myalgias, arthralgia, and rash. A small number of patients may develop haemorrhagic complications, such as gastrointestinal bleeding, subsequently leading to death. A small proportion of patients will develop severe dengue forms: dengue haemorrhagic fever (DHF) and dengue shock syndrome (DSS). DHF is characterized by the acute onset of high levels of fever, increased vascular permeability (plasma leakage), thrombocytopenia and hemorrhagic manifestations. DSS happens when there is fluid leakage into interstitial spaces that triggers hypovolemic shock which may lead to death (6,7).

The development of DENV vaccines or therapies is challenging. Primary infection with one of the four serotypes confers a protective immunity to the infecting serotype that lasts for a lifetime, and cross-protective immunity to the other three serotypes (heterotypic serotypes) that lasts for few months. Yet the subsequent infection with any of the heterotypic serotypes increases the chance of having DHF/DSS. A widely-accepted hypothesis of antibody-dependent enhancement (ADE) of this disease has been proposed to explain the increased risk in developing severe disease in patients with subsequent infection (8-10). Although most of the DHF/DSS cases occur in adults are associated with secondary infections, it is quite often that DHF/DSS cases are found in children, especially infants, during their primary dengue infection (6,11).

1.3 Dengue virus life cycle

The dengue virus life cycle involves two hosts, which are mosquito and human. When *Aedes* mosquito is blood-fed from humans that are infected with DENV on the viremic phase (a phase where the virus is found in the bloodstream (12)), the viruses are ingested into the mosquito and replicated in the epithelial lining of its midgut. The viruses then infect the salivary glands of the mosquito and eventually enter the saliva. This mosquito will remain infected for the rest of its life (6). During blood feeding of the infected mosquito on a new individual, DENV from the infected mosquito are injected into the blood stream of human. The viruses infect primarily the myeloid immune cells, including monocytes, dendritic cells, macrophages and hepatocytes located in liver, spleen or lung (13-15). DENV was also previously detected in lymphoid cells in the spleen and blood-clot samples based on the human autopsy study (15).

DENV infects host cells through the interaction between host cell's molecules and viral surface proteins. The putative host cell's molecules that have been proposed so

far included DC-SIGN (dendritic-cell-specific ICAM3-grabbing non-integrin), CD14 (cluster of differentiation 14), HSP70/HSP90 (heat shock protein 70/90 kDa), mannose receptor, heparin sulfate, macrophage Fc γ receptor and so on. However, no specific receptor/molecule that is necessary for DENV entry has been identified so far (16). DC-SIGN is a C-type lectin receptor expressed on the surface of dendritic cells and macrophages, and interacts with the mannose-type carbohydrate recognition domain of pathogens (17). This receptor interacts with the surface glycoprotein on the DENV, and mediates the DENV attachment to dendritic cells (18,19), but not through endocytosis (20). The Fc γ receptor has been shown to interact with the immune-complexes of mature or immature DENV, triggering the ADE phenomenon. It was suggested that in the subsequent infection, the heterotypic serotype of DENV induces the production of antibodies through the adaptive immune response. These antibodies form a complex with the DENV virion, which allows the viral entry via the Fc γ receptors (21,22) on the natural targets for the DENV such as macrophages and dendritic cells. This increases the viral load to the host and thus, enhances the disease.

The viral replication and assembly site of DENV occurs in the host cell's cytoplasm. The entry of viruses starts with the attachment of viral envelope (E) protein to the host cell's receptor and triggers the clathrin-mediated endocytosis. The acidification of endosomes induces conformational change on E proteins, which in turn allows the fusion of viral membrane to the endosomal membrane (23), with the subsequent release of the viral genome into the cytoplasm of the host cell. The positive-sense viral genomic RNA (gRNA) serves two roles in the cells: as the messenger RNA (mRNA) for the translation of viral proteins, and as the template for genome replication. Using the translational machinery from the host cells, the gRNA is translated into a single polyprotein, which embeds into the endoplasmic reticulum (ER) membrane. Consequently, this polypeptide is cleaved by the viral protease (NS3) or host proteases (eg. signalase and furin) to initiate the release of individual proteins. After cleavage, the structural protein, capsid (C), and non-structural (NS) proteins, NS3 and NS5, are released into the ER lumen, while other structural proteins, membrane precursor (prM) and envelope (E) proteins, and NS1, are released into cytoplasm (24,25).

The RNA genome is replicated in a replication complex, which is located in a double-membrane vesicle originating from the ER membrane (26). Then structural proteins, which form the viral envelope, and newly synthesized RNA genome are

assembled and budded into the ER lumen, forming the immature virus, which is generally non-infectious unless it complexes with the antibodies (10,27). Immature DENV has a diameter of about 60 nm with a ‘spiky’ morphology (Fig. 1.1A) (28). These immature viruses are transported to the trans-Golgi network (TGN), and in this compartment, the low pH (5.8-6.0) environment induces the rearrangement of the viral surface proteins. Furin, which is a host protease presented in the TGN, cleaves the precursor of membrane protein (prM) presented on the surface of viral particle. This generates infectious mature viruses, which has a diameter of 50 nm with a ‘smooth’ morphology (Fig. 1.1B) (24,25,28,29). The mature viruses are then secreted via exocytosis.

1.4 DENV virion structure

The mature DENV virion has a diameter of 50 nm. Three-dimensional image reconstruction from cryo-electron microscopy (cryo-EM) showed that DENV virion has an outer protein shell, a lipid bilayer and an inner nucleocapsid core, which is approximately 30 nm in diameter (30). The outer protein shell of DENV exhibits icosahedral organization, and is composed of 180 copies of E and M proteins. In immature viruses, prM and E proteins form the prM-E heterodimer, and this heterodimer further oligomerizes to form the trimeric prM-E spikes. The viral surface of immature virus consists of 60 copies of this trimeric prM-E spike, which give a “spiky” morphology to the virus (Fig. 1.1A) (31). In mature viruses, the whole viral particle surface is rearranged, where the precursor of prM is removed, and now M protein lies beneath E protein (32). The E proteins form homodimer, and viral surface in this stage is covered with 90 copies of this dimers, resulting in the observed “smooth” morphology (Fig. 1.1B).

Both E and M proteins contain the transmembrane domains which facilitates protein anchorage to lipid bilayer beneath this protein shell. The lipid bilayer has a thickness of 4.5 nm (Fig. 1.1C) (30). Beneath this bilayer, there is a shell formed by multiple copies of capsid (C) proteins, which interact with the genomic RNA, forming an approximately 30 nm nucleocapsid (Fig. 1.1C). Full-length capsid protein consists of a carboxyl-terminal hydrophobic signal peptide which is embedded into the ER membrane during translation. Then, the mature C protein is released by the proteolytic cleavage of viral protease (NS3). The C protein consists of 100 amino acids, and forms a homodimer in solution (33). It is highly basic, and its monomer consists of four

alpha-helices, with structurally disordered amino-terminus (residues 1-20) (34). The C protein homodimer has an asymmetric charge distribution, where one face is mainly formed by apolar residues and was proposed to interact with the lipid bilayer, while another face is mainly formed by basic residues and proposed to interact with the genomic RNA (34).

Figure 1.1 DENV virion structure. Figures (A) and (B) are reused from Perera *et. al.*, 2008, and figure (C) is reused from Kuhn *et. al.*, 2002, with the permission from Elsevier (25,30). (A) Immature DENV has a “spiky” morphology. (B) Mature DENV has a “smooth” morphology. (C) The cryo-EM density of central cross-section of DENV-2 cryo-EM reconstruction at 24 Å resolution. The *purple* and *blue* colored densities are the outer protein shell (E and M proteins). The density of lipid bilayer is colored in *green*. The nucleocapsid that is formed by capsid protein layer (colored in *yellow*) and genomic RNA (colored in *red*) lies beneath the lipid bilayer.

DENV genome is a single-stranded (ss), positive (+)-sense RNA with approximately 11 kb. It has a 5' type 1 cap (m7GpppAmp) (35). Since its gRNA is (+)-sense, this RNA can act as the mRNA in protein translation. Compared to the cellular mRNA, DENV gRNA does not have the poly-A tail at its 3' end (35). Its gRNA encodes a single open reading frame (ORF) flanked by 5' and 3' untranslated regions (UTRs). The 5' UTR is ~100 nucleotides long and conserves among different serotypes, and the 3' UTR is between 300 to 700 nucleotides long that carries a highly-conserved stretch of amino acid at its 3' ends. These UTRs have been proposed to be essential in promoting efficient viral RNA synthesis by adopting certain secondary structures (36,37).

Figure 1.2 Schematic diagrams of DENV (*top*) viral genome and (*bottom*) polyprotein. This figure is modified from Perera *et. al.*, 2008 and Li *et. al.*, 2014 (25,38). The viral genomic RNA contains 5' and 3' UTRs. After translation, the polyprotein is processed by the viral protease (NS3) (red arrow) and host cell proteases (blue arrow) to release the individual viral proteins.

The translation of DENV RNA can be achieved by both canonical cap-dependent and non-canonical cap-independent translation. The m7G cap is not essential for the translation of DENV RNA. It has been shown that the DENV replication and translation are not affected after the inhibition of the cap-binding protein eukaryotic initiation factor 4E (eIF4E) (39). Herein, eIF4E binds to the cap structure of the RNA and recruits other initiation factors to form the initiation complex for RNA translation to occur (40). In the non-canonical translation of DENV, unlike eukaryotic cells, DENV does not require an internal ribosome entry site (IRES) but requires the interaction between 5' and 3' UTRs of DENV RNA, suggesting a novel strategy utilized by DENV in its translational mechanism (39). The ORF is translated into a single polyprotein and further processed by the host and viral proteases into ten viral proteins shown in figure 1.2 (41). The structural proteins form the viral particles, while the NS proteins are found in the host cell but not in virions.

1.5 DENV genome replication

The electron microscopy and tomography showed that the DENV infection induces the alteration and proliferation on intracellular membrane to form several virus-induced membrane structures, including the convoluted membranes (CM), tubular structures (T) and vesicle packet (Vp) (42). The Vp, which is formed by two bilayer membranes derived from the ER, contains the replication complex (RC), is the replication site for the gRNA of DENV (*inset* of Fig. 1.3) (42-44). It is proposed that this architecture provides a confined and stable environment for the gRNA replication,

and protects the viral gRNA from being recognized by the host defense system (45,46). The RC comprises of the viral non-structural (NS) proteins, viral gRNA and several host proteins (42,45). All seven NS proteins are localized in the RC (42), where the NS3 and NS5 form the central catalytic core for the gRNA replication (Fig. 1.3). The NS3 comprises of activities of the helicase, nucleoside triphosphatase (NTPase) and RNA 5'-triphosphatase (RTPase) (47-49), while NS5 comprises of activities of the guanylyltransferase (GTase), methyltransferase (MTase) and RNA-dependent RNA polymerase (RdRp) (50-52). The membrane proteins NS2A, NS2B, NS4A and NS4B are possibly responsible for the membrane alteration (53) and/or association of the viral RC to the membrane (54,55). NS1, which is released into the ER lumen after protease cleavage, is shown to reside within the RC, even though the interaction between this protein and other NS proteins presented on the cytoplasmic face remains unclear (56).

Figure 1.3 Schematic diagram of the hypothetical model of DENV replication complex. This figure is modified from Muller and Young, 2008 (57). Invagination of the ER membrane forms vesicle packet (Vp) that contains the replication complex (RC). NS3 and NS5 form the RC catalytic center. NS2A, NS2B, NS4A and NS4B are embedded to the ER membrane. The protease domain of NS3 interacts with NS2B and the helicase domain interacts with NS4B. NS5 interacts only with NS3. NS1 is the only protein resides in ER lumen, and associates with NS4A and NS4B. (inset) The slice of EM tomogram shows an invaginated vesicle on ER membrane, and below shows a 3D surface of the vesicle. This figure is reused from Welsch *et al.*, 2009 with the permission of Elsevier (42).

The RNA synthesis of DENV is semi-conservative and asymmetric, where the (+) stranded RNA is predominantly synthesized over the negative (-) stranded RNA, and

the (-) stranded RNA is only detectable in double-stranded (ds) RNA form (35,58). Previous studies detected three forms of RNA species with different sedimentation coefficients in cells that were infected with DENV, including the genomic ss(+)-RNA (40 S), the double-stranded replicative form (RF) RNA (20 S) and the replicative intermediate (RI) (20-28 S), which is an intermediate containing growing ss(+)-RNA and the RF template (58,59). The overview of proposed scheme for DENV RNA replication (58,60,61) is shown in figure 1.4. The NS5 and NS3 were shown to convert the RF to the RI, and this conversion may involve one or more cellular proteins (59). After the translation of viral RNA, the host proteins and the viral proteins responsible for RNA replication form the replication complex. The RNA-dependent RNA polymerase (RdRp) domain of NS5 first synthesizes the dsRNA by producing the complementary (-) stranded RNA from ss(+)-RNA through semi-conservative replication. The NS3 helicase may be responsible for the unwinding of this dsRNA and then the new ss(+)-RNA starts to amplify on the (-) stranded RNA template, displacing the pre-existing (+) stranded RNA to form new RF RNA, which acts as the template for the next round of ss(+)-RNA synthesis.

Figure 1.4 The proposed mechanism of DENV RNA replication. This figure is modified from Klema *et al.*, 2015 (60). In the first cycle, the RdRp of NS5 synthesizes a new (-) stranded RNA from the (+) stranded genomic RNA and forms the double stranded RNA, which is termed as the replicative form (RF). In the next cycle, the helicase of NS3 may unwind the RF, and a new (+) stranded RNA is replicated on the RF, forming the replicative intermediate (RI). After finishing this cycle, one (+) stranded RNA and one RF are produced, and the RF is recycled and start another cycle of replication.

The newly synthesized ss(+)-RNA is then subjected to RNA capping. Eukaryotic cells in general perform the capping mechanism inside their nuclei. Since the capping mechanism of flaviviruses is in the cytoplasm, these viruses encode their own capping enzymes (51). In one of the flaviviruses, WNV, it is shown that the cap is only present at the 5'-end of gRNA, but not in the dsRNA (35). The overview of RNA capping mechanism (60,62) is shown in figure 1.5. The first step is believed to be carried out by the NS3 protein which has the RTPase activity in its helicase domain (48). In this step, the γ -phosphate at the 5'-end of RNA is hydrolyzed to create a diphosphate RNA (pp-RNA) and an inorganic phosphate (P_i) is released. Afterwards, the GTase in the MTase of NS5 removes the pyrophosphate (PP_i) from the guanosine triphosphate (GTP) to form guanosine monophosphate (GMP), and then transfers this GMP to the pp-RNA to form the Gppp-RNA. This Gppp-RNA is methylated by the MTase of NS5, by transferring a methyl group from *S*-adenosyl-methionine (SAM) onto the N-7 position of the guanine to generate the product cap-0 RNA (m^7 Gppp-RNA) and the by-product *S*-adenosyl-homocysteine (SAH). Second methyl group is later on added to the ribose-2'-O positions of the first nucleotide of cap-0 RNA by the same MTase, creating the final RNA product, cap-1 RNA (m^7 Gppp- $N_{m2'O}$ -RNA).

Figure 1.5 The proposed mechanism of DENV RNA capping mechanism. This figure is modified from Klema *et al.*, 2015 (60). The γ -phosphate of 5'-end of newly synthesized (+) stranded RNA is believed to be hydrolyzed by the RTase in the helicase of NS3. The GTP is activated by the GTase in the MTase of NS5. The GTase transfers the guanine cap to the pp-A1-RNA and forms Gopp-A1-RNA. Then the MTase of NS5 transfers a methyl group from SAM to G₀, forming cap-0 RNA and releases SAH as the by-product. Next, the methyl group is transferred by the MTase to A₁ and forms the final RNA product, cap-1 RNA.

1.6 RNA cyclization in DENV genome replication

The 5' and 3' UTRs of DENV ss(+)-RNA contain a few sequences that are necessary for RNA replication (36,63). Several secondary structures are identified on both the UTRs of DENV RNA, including stem-loop A (SLA), stem-loop B (SLB) and capsid region hairpin (cHP) located at the 5' UTR, and 3' terminal stem-loop (3'SL) located at the 3'UTR of DENV RNA (Fig. 1.6A) (37). To discriminate its own viral RNA from the host cellular RNA, the RdRp of DENV specifically recognizes a 5' RNA element named SLA (36), but not the 3' end of RNA (64). Mutations within the SLA region showed that the SLA is essential for the recruitment of viral RdRp, and the disruption of its stem loop structure affects the RNA replication. However, if the stem loop structure is preserved, even with the non-viral sequence, the RNA synthesis is not affected (36). The 3' UTR of viral RNA was shown to be important for RNA replication, but not RNA translation (63). It has also been shown that DENV-1 NS3 protein binds to the 3'SL of DENV RNA (65).

Figure 1.6 Schematic diagram of DENV RNA cyclization. This figure is adapted from Villordo *et al.*, 2009 (37). (A) Structural elements located in the DENV genomic RNA. (B) The cyclization of genomic RNA is achieved by two complementary sequences: 5'-3' UARs and 5'-3' CSs.

From the image obtained using atomic force microscopy (AFM), the DENV RNA forms a circular conformation (66). Two sets of complementary sequences are identified at 5' and 3' end of viral RNA: 5'-3' CS (complementary sequence) and 5'-3' UAR (upstream AUG region), which are responsible for this cyclization of the RNA (Fig. 1.6B). With these two sequences located at 5' and 3' UTR, the 11 kb RNA genome forms the long-range RNA-RNA contacts and cyclizes to promote RNA synthesis (66). It has been shown that the complementation of these sequences but not the nucleotide sequence itself is important for enhancing RNA synthesis (36). The 3' SL at the 3' UTR can repress RNA synthesis, and this silencing effect is released after cyclization of the RNA (64). During hybridization of complementary sequences, the 3' SL may undergo certain conformational changes, where the 3' UAR is overlapped with the 3'SL sequence (Fig. 1.6A), and thus, releasing the silencing effect of the 3' SL (37,64). This RNA cyclization is proposed to provide several advantages to DENV, including: ensuring only full-length RNA amplification, coordinating RNA replication and translation through overlapped signal from 5' and 3' UTRs, stabilizing the RNA, recruiting the replication complex to the correct site, and controlling the replication of (-) stranded RNA (36).

A model for (-) stranded RNA synthesis has been proposed based on these observations (37). The interaction between DENV NS5 and viral RNA is mainly achieved by the RdRp domain (64). In the proposed mechanism, the viral RdRp recognizes the SLA, which is located at the 5' end of the viral RNA. The viral RNA adopts a circular conformation through its 5'-3' CS and 5'-3' UAR complementary sequences, which in turn allows the close proximity between the 3' end of RNA and the RdRp, and initiates the (-) stranded RNA synthesis. Since the template channel shown in the crystallographic structure of DENV RdRp (PDB ID: 2J7U (67)) can only accommodate ssRNA, it is likely that the NS3 helicase unwinds the 3'SL before the replication started (37).

1.7 Non-structural protein 5 (NS5)

NS5 is the largest proteins in DENV, with about 900 amino acids. It is a two-domain protein, and comprises of the RNA-dependent RNA polymerase (RdRp) domain and methyltransferase (MTase) domain. RdRp is responsible to the synthesis of the new strand of viral RNA, and the MTase carries out three sequential enzymatic reactions that are related to the capping mechanism. The 5' cap of viral RNA can

provide a protection to the RNA against other cellular exonuclease, and also improve the recognition of RNA to the eIF4E, which in turn ensures efficient translation of RNA (62). The physical linkage between two domains suggests that the viral RNA replication and capping mechanism are coupled, and it was shown that the MTase domain has a stimulation effect on the RdRp activity in DENV (68).

DENV-2 NS5 was shown to interact with NS3 *in vivo*. However, in the same study, these two proteins predominantly localized in different cell compartments, where the NS5 was predominantly found in the nucleus, while the NS3 resides in the cytoplasm (69). Further analysis showed that DENV NS5 exists in multiple forms due to the differential phosphorylation on serine residues at multiple sites. The hyperphosphorylated NS5 was predominantly found in the nucleus, while only the cytoplasmic form of NS5 was co-immunoprecipitated with NS3 (69). Later on, two functional nuclear localization sequences (NLS) are identified on DENV-2 NS5, which are located at residues 320-368 and 369-405, termed bNLS and a/bNLS, respectively (70,71), and this nuclear localization is mainly contributed by the a/bNLS (72). Within this region, the residues K371 and K372 from DENV-2 NS5 are identified as the critical residues for nuclear transport (71). Comparing the nuclear localization of the NS5 across all four DENV serotypes, the NS5 of DENV-1 and DENV-4 predominantly localize in the nucleus, while DENV-2 and DENV-3 predominantly localize in the cytoplasm (73,74). This serotype-specific difference in nuclear localization might be due to the absence of functional a/bNLS in the NS5 of DENV-1 and DENV-4 (72,73).

In addition to its roles in the viral replication, the NS5 was shown to help DENV evade the innate antiviral response. STAT2 is one of the key players that is involved in the interferon signaling pathway to eradicate pathogens. The NS5 inhibits interferon- α signaling pathway specifically, and the RdRp domain alone is sufficient to trigger this inhibition (75). It has been shown that the NS5 binds to the STAT2 specifically and inhibits STAT2 phosphorylation (75,76). The NS5 can be expressed as a precursor, which contains the full-length NS5 and the extreme N-terminus located upstream of the NS5. This NS5 precursor is able to reduce the STAT2 expression level by directing it to a degradation pathway (76). This association and degradation of DENV NS5 and STAT2 are species-specific to the human STAT2, but not murine STAT2 (77).

1.7.1 Methyltransferase (MTase)

The MTase domain in NS5 is a multi-functional enzyme. It carries out RNA guanylyltransferase (GTase), guanine-N7 methyltransferase (N7-MTase) and nucleoside-2'-O-methyltransferase (2'-O-MTase) activities. The structure of MTase of DENV can be sub-divided into three subdomains: N-terminal GTP-binding subdomain (residues 7-54), catalytic core subdomain (residues 55-222), and C-terminal subdomain (residues 223-267). The N-terminal GTP-binding subdomain contains a GTP-binding site which provides GMP for cap formation. It is linked to the core subdomain via a positively charged protein stretch (50). The core subdomain has the Ado-Met (SAM) binding site and the MTase catalytic site. The C-terminal subdomain's function is unclear.

After the hydrolysis of γ -phosphate at the 5'-end of RNA by the NS3, the GTase activity of MTase removes the pyrophosphate from the GTP and transfers GMP to the ppRNA. The crystallographic structure of MTase-GTP-SAH complex (PDB ID: 2P1D (51)) showed that the GTP bound to the MTase through hydrogen bonding with L17, N18, L20 and S150, and staking interaction with F25. These four residues are highly conserved among flaviviruses (Fig. 3.1.2). From the mutational studies, four of them, N18, F25, K29 and S150, are identified as the critical residues for GTP binding (51).

The core subdomain of the MTase adopts an AdoMet-dependent MTase fold, where it folds into seven β -strands surrounded by four α -helices. It has a cavity which has been discussed to be the active site for catalysis and substrate binding (51,78). Structural analysis of the crystallographic structures of the DENV-3 MTase-SAM complex (PDB ID: 3P9Z (79)) and the DENV-2 MTase-SAH complex (PDB ID: 1L9K (51)) reveals a cavity located beside the SAM/SAH binding pocket. This cavity is unique to the flaviviruses and important for the cap synthesis as well as virus replication in the host cell and thus, it has been proposed as one of the drug targets (79). The binding of SAM/SAH to the MTase is mainly through the hydrogen bonding and van der Waals interactions (51), and residues in the SAM/SAH binding sites are highly conserved among flaviviruses (Fig. 3.1.2). Mutation on these residues abolished or reduced the N-7 and 2'-O MTase activity. The MTase catalyzes the methylation at two different positions on the cap of RNA (G₀pppA₁-RNA) in a sequential manner. Firstly, the MTase methylates the N7 position of the G₀ and then to the ribose 2'-O of the A₁. Both N-7 and 2'-O methylations acquire the methyl group from single SAM-

binding pocket in the MTase (80). Four highly conserved residues K61, D146, K181 and E217 located at the catalytic site form the catalytic tetrad K-D-K-E that are responsible for the N-7 and 2'-O MTase activities (51,81). It was shown in the West Nile virus study that the catalytic tetrad K-D-K-E is essential for 2'-O MTase activity, whereas only D146 is essential for N-7 MTase activity (81). One study showed that inactivation of the 2'-O methylation reduced DENV replication, while inactivation of both the N-7 and 2'-O methylation in DENV abolished viral replication. This signifies that the N-7 methylation is more important than 2'-O methylation in viral replication and thus, the N-7 methylation is proposed as a potential drug target (80). The study on YFV NS5 revealed a highly-conserved residue: S56, amongst flaviviruses. Furthermore, the phosphorylation of S56 is essential for the 2'-O MTase activity and viral replication (82).

Recently, the crystallographic structure of DENV-3 NS5₆₋₈₉₅ solved together with cap-0 viral RNA (^{m7}G_{0ppp}A₁G₂U₃U₄G₅U₆U₇) and SAH (PDB ID: 5DTO (83)) provides more insights to the MTase-RNA interaction (Fig. 1.7). The guanosine ring of G₀ is interacting with the amino acids N18 and L20 of the MTase domain, and this interaction is enhanced by the stacking interaction between N-7 methyl group of G₀ and F25 of MTase. These interactions are same as the interactions found in the crystallographic structure of MTase-GTP-SAH complex (51). The 2'-O-ribose of nucleotide A₁ is positioned close to the SAH (Fig. 1.7). The analyzed crystal was obtained by soaking the native NS5 crystal with the cap-0 viral RNA, and no reorientation between MTase and RdRp domains were identified upon the binding of capped RNA, suggesting a preformed RNA binding groove (83,84). The A₁ and G₂ are strictly conserved among flaviviruses. Based on this crystallographic structure, the modeling with other nucleotides had been performed, and they showed that the replacement of these two nucleotides with other nucleotides led to unfavorable contacts between their bases and the binding pocket of the MTase (83), indicating the specific recognition of the viral RNA by the MTase. However, despite high similarity of MTase and RdRp domains among the viruses from flavivirus genus, it has been shown that the chimeric DENV, which contains WNV MTase/RdRp or full-length NS5, were non-replicative. Thus, indicating the capping mechanism difference among flaviviruses (80).

Figure 1.7 The MTase domain taken from the crystallographic structure of DENV-3 NS5₆₈₉₅-RNA co-crystal (PDB ID: 5DYO) (83). The structure is shown as (left) *cartoon* and (right) *surface representation*. The MTase is colored based on its subdomains, where the N-terminal GTP-binding domain is in *dark blue*, the catalytic core domain is in *blue* and the C-terminal domain is in *light blue*. The G₀pppA₁-RNA and SAH are represented as *stick*. The nucleotides G₀ and A₁ from the capped RNA are colored in *magenta*.

1.7.2 RNA-dependent RNA polymerase (RdRp)

The RdRp domain of NS5 carries out the replication process of the viral genome. Due to the lack of proofreading exonuclease activity in the RdRp, RNA viruses generally exhibit a high mutation rate in their viral genome replication (85). As seen from the DENV-3 RdRp crystallographic structure (residues 265-900; PDB ID: 4C11 (86)), it adopts a similar structure like the other template-directed polymerases, and has a U-shaped right-hand structure with three subdomains: fingers, thumb and palm (Fig. 1.8) (50,87). RdRp has a unique structure compared to other types of polymerases, where it has a “fingertips” (*yellow* extension from the *red* fingers subdomain in Fig. 1.8) extending from the fingers subdomain that interacts with the thumb subdomain and encloses the active site. This interaction restricts the movement of these two subdomains and hence, no significant conformational change is observed upon binding of the substrates (88). Six conserved motifs are identified within the DENV RdRp, and they are important for deoxy-nucleoside triphosphate (dNTP), RNA and metal ions binding, and catalysis (50). The fingers and thumb subdomains are responsible for RNA primer and template binding, while the palm subdomain is important in positioning the primer, divalent cations, templates and incoming nucleotides (87).

Figure 1.8 The crystallographic structure of DENV-3 NS5₂₆₅₋₉₀₀ (RdRp; PDB ID: 4C11) (86). The structure is shown as (left) cartoon and (right) surface representation. The RdRp is colored based on its subdomains, where the fingers subdomain is in red, palm subdomain is in green and thumb subdomain is in violet. The bNLS peptide is colored in yellow. The GDD catalytic active site is labeled as black stick.

Palm subdomain is the most structurally conserved among the polymerases. The overall architecture of catalytic site is preserved during evolution. Four (Motif A–D) of six motifs are located at the palm subdomain (87,89). The conserved motif C contains the Gly-Asp-Asp (GDD) catalytic active site (residues 662–664) (black stick in Fig. 1.8). The motif A, B and D are responsible for the binding and positioning of dNTPs, metal ions and RNA template. The conserved Asp residues located at motif A (Asp-533) and C (Asp-664) mediate the Mg^{2+} ion binding (87).

Together with the thumb subdomain, the fingers subdomain shapes a tunnel that directs the incoming RNA to the active site. Two protein stretches in this subdomain were shown to contain the functional nuclear localization sequences, named bNLS and a/bNLS (70,71,90). The bNLS domain (residues 316–368), which is located at the “fingertips” of RdRp, has a helix-turn-helix motif that lies on top of the thumb subdomain (yellow cartoon in Fig. 1.8). It was shown to interact with importin β and NS3 helicase (71,90,91). Truncation of this NLS domain leads to destabilization of the RdRp structure. This subdomain contains one conserved motif, which is motif F. This motif consists of positively charged residues that are required to mediate the interaction of the RdRp with incoming dNTP (50).

The thumb subdomain is the most structurally variable among the known polymerase. Compared to other viral RdRps, the RdRps of flaviviruses usually have larger thumb subdomain that narrow down the template-binding channels. This allows

a selectivity to the single-stranded RNA (ssRNA) and dNTPs in the *de novo* initiation, which initiate RNA synthesis by adding a second NTP to the 3'-hydroxyl of the first NTP without primer (88). The thumb subdomain contains two conserved motifs: motif E and priming loop. Together with the N-terminal loop of the finger subdomain, the priming loop (residues 792-804) forms the upper part of template tunnel and may play a role in regulating the entry or exit of the template to the RdRp (67). The residues His-712, His-714 and Cys-728 located at motif E interact with the Zn^{2+} ion. This ion is suggested to contribute to the stability of the region near motif E, and since it is located near the nucleotide binding site, it may have a role in RNA replication. The template tunnel of the DENV RdRp allows only the binding of ssRNA. Therefore, the loops within the template tunnel are likely to adopt different conformations to accommodate the newly synthesized duplex RNA (67).

1.7.3 10-residue linker

The ten residues of DENV NS5 (residues 263-272) form the flexible linker region connecting MTase and RdRp. Although the number of linker residues is conserved among the flaviviruses, their linker sequences are poorly conserved (Fig. 3.1.1). This linker physically connects two enzymes that catalyze various events in the viral genome replication, suggesting the coupling mechanism between the RNA replication and capping mechanism. In the absence of the RdRp, the MTase activity was reduced by 2-fold as compared to the full-length NS5 (92). Furthermore, without the MTase domain, RNA synthesis was affected (68) and thus, pointing out the importance of this coupling effect. However, it was also reported that the MTase and the RdRp activities function independently (93).

The DENV-3 RdRp (residues 265-900; PDB ID: 4C11) with the extension of seven residues at N-terminus of RdRp was shown to stabilize the RdRp and improve the efficiency of RNA synthesis compared to the RdRp that spans residues 272-900 (86). Mutational studies on the linker residues reveal that these residues improve the catalytic properties of the RdRp by increasing the turnover of RNA and NTP, but not to the binding affinity between substrates and RdRp (86). The recently solved DENV-3 NS₅₆₋₈₉₅ crystallographic structure (PDB ID: 4V0R) has a fully resolved linker region (²⁶³HVNAEPETPN²⁷²), where the residues ²⁶³HVNA²⁶⁶ form a short ₃₁₀-helix (84). This short ₃₁₀-helix is missing in the JEV NS5 structure (PDB ID: 4K6M) (94). It was proposed that the ₃₁₀-helix in DENV and/or the highly conserved ²⁶⁰GTR²⁶² may

function as a swivel to open the interface between two domains and provide structural transition to the NS5 (84,94). The residues E267 and E269 are conserved among all four DENV serotypes and form the polar contacts with the MTase and the RdRp (84,86).

Mutations on these two strictly conserved linker residues in DENV-3 NS5 affect the RdRp activity, where the E269A mutant has reduced activity while the E267A mutant has an increased RdRp activity compared to the wild-type. However, the MTase activity of both mutations showed a 30-40% reduction, suggesting a minor allosteric effect coming from the RdRp (84). Furthermore, mutational study on another strictly conserved linker residue V264 in DENV-3 NS5 (or I265 in DENV-2) that leads to a change in its linker flexibility, showed the importance of maintaining optimal conformation flexibility in NS5 to achieve optimal viral replication (95).

1.8 Interaction of NS3 and NS5

The non-structural proteins, NS3 and NS5, form the catalytic core in the replication complex (RC), and carry out two important events in the viral replication cycle: viral genome replication and genome capping. The interaction of DENV NS3 and NS5 were firstly characterized *in vivo* and *in vitro* based on biochemical methods (69). From the same study, the interaction between these two proteins was proposed to be regulated by the phosphorylation of NS5. The interaction between NS3 and NS5 was also shown in yeast two hybrid assays (90). The interacting sites are further mapped to the bNLS region (residues 320 to 368) of the NS5 and residues 303-618 in NS3 (90), and this interaction is specifically to the bNLS region, not a/bNLS (71). In addition, bNLS region of NS5 not only interacts with NS3 but also interacts with importin- β , which is a protein that is involved in protein transportation into the nucleus. Therefore, it is postulated that NS3 may have an additional role in ensuring that NS5 remains in the cytoplasm (90). The interaction site on the NS3 is further mapped to residues 566-585 based on the ELISA outcome (91).

The importance of this interaction is also shown in a mutagenesis study, where mutating the bNLS region of recombinant NS5 has little effect on its nuclear localization. However, the mutation on this region abolishes viral replication when the mutation is introduced into the viral genome (71). NS5 was shown to have the stimulatory effect on the NTPase activity of NS3 (65). By adding DENV-1 NS5 into the assay, the NTPase activity of NS3 was increased. The stimulation of the NTPase

activity by NS5 is further confirmed using NS3 and NS5 from DENV-2. A surface plasmon resonance (SPR) assay conducted using full-length NS5, the helicase domain of NS3 and full-length NS3 reveals that the binding affinity of full-length NS5 with full-length NS3 is higher than the helicase domain alone, indicating that the protease domain also participated in the interaction with the NS5 in order to enhance the NS3-NS5 interaction (96). This interaction is supported by a previous mutagenesis study on the MTase of NS5, where it suggests the MTase interacts with the NS3 protease domain (97). Later on, a robust *in vitro* assay used to study the interaction between NS3 and NS5, AlphaScreen assay, was developed, and proving that both protease and helicase associate with NS5 (98).

1.9 Non-structural protein 3 (NS3)

DENV NS3 (residues 1-618) is a multifunctional protein that is composed of two functional enzymatic domains: the protease and helicase domains. The protease domain is involved in releasing viral proteins from the polyprotein, while the helicase participates in RNA replication (2). To date, three crystallographic structures of DENV-4 NS2B₁₈NS3 were solved, and have identified two conformations. The 2VBC and 2WZQ structures adopt the same conformation, termed as conformation I (99,100), while the 2WHX adopts conformation II (99), whereby its protease domain is rotated 161° with respect to the helicase domain compared to conformation I (Fig. 1.9). From these two conformations, only conformation II allows the binding of ADP-Mn²⁺ (Fig. 1.9). When the ADP is computationally docked into the NTP binding pocket of conformation I, steric hindrance occurs between the ADP and the protease domain (99).

The protease and the helicase of NS3 from DENV are connected via a 10/11-residue linker. Similar to DENV NS5, the linker residues of DENV NS3 has low conservation. It is generally agreed that the optimal helicase activity requires the protease domain (25,101). In the absence of the protease, ATP binding affinity, ATPase activity and unwinding activity in the helicase are reduced (99,100). The mutational study on linker region of the DENV-4 NS3 by increasing its flexibility showed that the ATPase and helicase activities were significantly reduced. It is speculated that the ATP binding affinity to the NS3 will be affected by the linker flexibility, which eventually lead to a decrease in the enzymatic activities (99).

Figure 1.9 Crystallographic structures of DENV-4 NS2B₁₈NS3 adopted conformation I (PDB ID: 2VBC; (100)) and conformation II (PDB ID: 2WHX; (99)). The structures are represented as *cartoon*. (Left) The protease of the 2VBC structure is colored in *wheat*, the helicase is in *red* and the NS2B₄₉₋₆₆ is in *light orange*. (Middle) The protease of the 2WHX structure is colored in *sky blue*, the helicase is in *blue* and the NS2B₄₉₋₆₆ is in *light blue*. The ADP is represented as *green stick*. (Right) Superimposition of both conformations reveals a 161° rotation of the protease with respect to the helicase.

Apart from its roles in virus replication, the NS3 is involved in autophagy by interacting with the host cell proteins. Both DENV-2 and WNV NS3 have been shown to induce apoptosis in the host cells (102,103), and from WNV NS3 study, NS3 alone is sufficient to activate the caspase-8 activity that initiates apoptotic signaling (102). In addition, NS3 may work together with the host cell fatty acid biosynthetic pathway to form the replication complexes. The NS3 was shown to interact with the fatty acid synthase (FASN) and localize FASN to the DENV replication site (104). This process requires a small GTPase, Rab18, which normally acts as a molecular switch in the regulation of membrane trafficking (105).

1.9.1 Protease

The protease domain of the NS3 is a chymotrypsin-like serine protease that possesses both *cis* and *trans* proteolytic activities. It contributes to the release of individual viral proteins from the polyprotein after translation. It cleaves the polyprotein preferentially after adjacent basic residues (106). The protease domain requires NS2B as a cofactor to fold and function properly. NS2B is a membrane protein that consists of a N-terminal transmembrane (TM) domain, a central domain and a C-terminal TM domain. The central domain, which spans from residues 49-96,

forms a “belt-like” structure that wraps the protease domain (*green cartoon* structure in Fig. 1.10), and is essential for the protease domain to function as an active enzyme (107). Without this cofactor, the protease is not soluble. In order to obtain a soluble and stable construct for structural studies, this NS2B fragment is physically linked to the protease or full-length NS3 with a Gly₄-Ser-Gly₄ linker (99,100,108-110). Subsequently, the protease with an unlinked NS2B was generated for NMR experiments and is catalytically active, providing an improved NMR spectrum that benefits future drug screening (111).

Figure 1.10 The protease domain of DENV-3 NS3 in complex with the NS2B₅₀₋₉₅ and the inhibitor (PDB ID: 3U1I), which adopted a closed conformation (109). The structure is shown as (*left*) cartoon and (*right*) surface representation. The protease is colored in *slate blue*, and the NS2B₅₀₋₉₅ is in *green*. The peptide-based inhibitor, benzoyl-norleucine-Lys-Arg-Arg-aldehyde is colored in *magenta*.

Besides the protease expressed together with 47 residues from NS2B, it was reported that the 18 residues of NS2B (residues 49-66) is sufficient to generate a stable protease domain, but this protease is enzymatically inactive (100). This signifies the importance of NS2B in regulating the enzymatic activity of the protease. From the crystallographic structures, as well as the NMR data, the C-terminal central domain of NS2B is flexible, and the folding states of this region determine the “open” or “closed” conformation of the protease (108,109,111). The protease domain is capable of being transactivated by providing the NS2B externally to the protease *in vitro*. It is also shown that the DENV-2 protease can interact and be transactivated by the NS2B from

DENV-1, -3, -4 and WNV, showing the conservation of the activation mechanism across flaviviruses (112).

1.9.2 Helicase

Helicase domain of the NS3 is required in the viral RNA replication, where it possesses RNA unwinding activity, NTPase activity and RTPase activity (47,48,113). The helicase consists of three subdomains (Fig. 1.11), where two conserved domains are required for the NTP hydrolysis, and a unique C-terminal domain that may play a role in recognizing virus-specific RNA and protein (114). The RNA binding tunnel is located at the middle of all three subdomains (Fig. 1.11). The C-terminal domain was shown to be essential for RTPase activity, but the phosphodiester bond hydrolysis is catalyzed in the catalytic core by utilizing the Walker B motif, which is involved in Mg^{2+} coordination (114). NS3 interacts with NS5 through subdomain III of the helicase (71), and it is further mapped to residues 566-585 in subdomain III (91).

Figure 1.11 The helicase domain of DENV-3 NS3 in complex with the ssRNA (12-mer) and the AMPPNP (PDB ID: 2UJV) (115). The structure is shown as (left) cartoon and (right) surface representation. The helicase is colored as yellow for the subdomain I, orange for the subdomain II and red for the subdomain III. The ssRNA (5'-AGACUAACAACU-3') is represented as cyan stick and the AMPPNP is showed as green stick. The blue arrow indicates the proposed "helix opener".

The RTPase and NTPase activities share the same active site in the helicase (47). The RTPase activity in helicase is required for the first reaction in the viral RNA cap synthesis, where it dephosphorylates the 5' end of viral genomic RNA to allow the MTase in NS5 to add the guanine cap to the viral RNA. The interaction between the RNA and the helicase is electrostatic, and was shown to be sequence-independent in

biochemical and crystallographic studies (47). The biochemical study detected the unwinding activity when utilizing non-DENV dsRNA sequence as the RNA substrate (47), while the crystallographic studies reported two crystallographic structures that co-crystallize with two different ssRNAs (PDB ID: 2JLU and 2JLW) (100). Based on the existing crystallographic structures, the dsRNA is proposed to be unwound by a “helix opener”, which is a β -hairpin from the subdomain II (*blue arrow* in Fig. 1.11). This β -hairpin disrupts the dsRNA and stabilizes the unwound ssRNA. The positively-charged region between subdomain 2 and 3 resides close to this hairpin and it may interact with the 3' overhang sequence of dsRNA to unwind RNA (100,116). The energy utilized in these unwinding and translocating processes may be supplied by the NTP hydrolysis, although the exact mechanism that couples these events remains unclear (116).

1.10 NS3 and NS5 as a target for DENV vaccine or antiviral drug development

As mentioned earlier, NS3 and NS5 form the central catalytic core in the RC. They possess the important enzymatic activities in virus replication. Therefore, they have been targeted for many years in DENV vaccine and antiviral drug development.

The MTase of NS5 is an enzyme involved in the capping mechanism of the DENV genomic RNA. Although the capping enzymes from eukaryotes and viruses are similar, there are still significant differences in their gene organization and catalytic mechanism, which make the RNA capping enzyme an attractive drug target. Furthermore, although different RNA viruses utilize different mechanism in their RNA capping, they lead to a similar RNA cap structure, suggesting the evolutionary importance in keeping this cap structure to protect the RNA (62). In the MTase, a specific cavity beside the SAM binding pocket was identified as one of the potential drug target. The SAH/SAM analog was designed to be selectively targeted to this specific cavity using a structure-guided approach (79). In another high throughput screening using the fragment-based experimental approach, seven fragment hits were identified, and three novel binding sites were identified in the MTase, where the amino acids presented in these binding sites interact with the fragments (117).

The RdRp of NS5 plays a role in the genomic RNA replication. Due to the lack of the proofreading mechanism in RdRp, high mutation rates lead to the development of drug-resistant viruses which eventually hinder drug development for the flaviviruses (118). However, this lack of proofreading mechanism makes RdRp a good drug target.

The nucleoside analog inhibitors have been designed to target the RdRp, in which they compete with natural substrates and incorporate into the viral RNA. For example, the ribavirin, which is a synthetic nucleotide analog, can be incorporated into the viral RNA during RNA replication to increase mutation rates and eventually lead to “lethal mutagenesis” (119). Another type of nucleoside analog, which triggers the RNA chain termination, was developed for the hepatitis C virus (HCV), a closely related virus to flavivirus (96,120). Besides the inhibitors targeting the RNA synthesis, an allosteric inhibitor that binds to RdRp pocket has also been considered in order to lock the conformational changes that are required for the function of the polymerase (121).

The protease of NS3 is responsible for the release of the individual viral proteins from a polypeptide. Although several inhibitors that mimic the substrate of protease have been designed, none of them has been tested in the clinical stage. Drug development based on the protease is challenging due to the shallow and solvent-exposed substrate binding pocket, which reduces the affinity between inhibitors and protease (38). Most of the reported peptidic inhibitors have at least one positively-charged residue that reduces cell permeability (121). The protease requires NS2B for its proper folding and activity. The cross-transactivation of the DENV-2 protease by the NS2B from DENV-1, -3, -4 and WNV suggests a potential drug target in the allosteric inhibitor development. It was shown that the human monoclonal antibody 3F10 could disrupt the interaction between the protease and NS2B, and mammalian cells infected with 3F10 were defected in the DENV replication (112). The library screening and *in silico* docking have been utilized to identify the potential leads, but most of the compounds did not show activity in cell-based assays (121).

The helicase of NS3 is involved in both RNA replication and capping mechanism. The lack of substrate specificity in the NTP and RNA binding pocket challenges the design of an inhibitor targeting the NTP-helicase and RNA-helicase interactions (121). Based on the computational docking into the RNA binding pocket of WNV helicase, an inhibitor, ivermectin, which was also shown to inhibit the protease, was identified. When this inhibitor tested in cell culture, it cross-reacted with DENV helicase, albeit with less efficiency (122).

1.11 Capsid protein

The genome of DENV encodes three structural proteins that form the virion, including envelope (E), membrane (M) and capsid (C) proteins. The cryo-EM showed

the genomic RNA (gRNA) of DENV is encapsidated by an outer protein shell formed by E and M proteins, a lipid bilayer, and an inner capsid shell (Fig. 1.1C) (30). Multiple copies of C protein interact with gRNA to form the nucleocapsid core. The mature C protein (residues 1-100) is a highly basic protein, and exists as a dimer in solution, which may be the building block for nucleocapsid formation (33,34). The immature C protein (residues 1-114) contains a hydrophobic C-terminus that functions as a signal peptide for the translocation of the polypeptide into ER. This C-terminal region is then cleaved by the NS3 protease, and the mature C protein is released into the cytoplasm (2).

The atomic resolution structure of DENV-2 C protein was obtained in the solution using nuclear magnetic resonance (NMR). Due to the labile conformation at the N-terminus of C protein, the first 20 residues are not included in the solved DENV capsid structure (34). The C proteins form the dimer, and each monomer consists of four α -helices and a disordered N-terminus. The C dimer has an asymmetrical charge distribution, where the hydrophobic side is proposed to interact with the lipid bilayer, while the charged side may interact with the viral RNA (34). It is reported that the purified DENV C proteins can undergo *in vitro* assembly using single-stranded DNA and form the nucleocapsid-like particles. This assembly reaction is independent of the sequence and length of DNA, suggesting the driving force of this reaction is the neutralization of this positively-charged C dimer (123).

DENV C proteins were proposed to interact with the lipid droplet (LD), which is an ER-derived organelle inside the host cell cytoplasm, via its N-terminal region, providing a scaffold for the encapsidation of viral RNA (124). Perilipin 3, which is a protein in the LD, is identified as the protein responsible for the interaction between the C protein and the LD. Besides the LD, the C protein also interacts specifically with very low-density lipoprotein (VLDL), but not low-density lipoprotein (LDL), indicating that there are one or more proteins present on the VLDL but not LDL, that interact with the C protein (125). The apolipoprotein E (APOE), which is a protein that is normally found in the VLDL but present in very low amounts in the LDL, adopts a similar fold as the perilipin 3. The APOE is proposed to interact with the C protein and mediate the interaction between the C protein and VLDL (126).

The N-terminus of the C protein is shown to be intrinsically disordered in solution. In a peptide study, the pep14-23, which contains residues from the N-terminal C

protein, folds into a α -helical conformation in the presence of phospholipids (127). It is reported that this peptide can disrupt the interaction between the C protein and the VLDL or the LD, which is important to virus replication, providing a new therapeutic approach to the DENV (126,128).

1.12 Small-angle X-ray scattering (SAXS)

Small-angle X-ray scattering (SAXS) is a low-resolution technique that characterizes the biological macromolecules such as proteins in solution in a structural and dynamic context. It can provide the structural information at different levels. In the absence of an atomic resolution structure, studying these proteins in solution can provide the first structural insight into its molecular shape. If a high-resolution structure is available, SAXS can provide further insights into the macromolecule structure in solution, including single proteins, multi-domain proteins, protein assemblies, protein-nucleic acid complexes, as well as protein-lipid complexes.

Macromolecule X-ray crystallography and nuclear magnetic resonance (NMR) are two common techniques that provide the macromolecule structural information at atomic resolution. X-ray crystallography provides the most number of the atomic resolution structures in Protein Data Bank (PDB) (129). Both SAXS and X-ray crystallography share a similar concept, in which the incident X-ray beam irradiates a sample and the scattered beam is detected. A fundamental difference between these two techniques is the requirement of the relative molecular orientation during data collection (130). SAXS averages the signal from all orientations of the sample molecules and thus, the sample can be analyzed in solution and no additional sample preparation step is required. In contrast, X-ray crystallography requires a highly-organized crystal lattice, and this organization provides discrete diffraction maxima that retain information of specific orientations of the sample molecules (130). However, the flexible regions inside the proteins may hinder them from the proper packing in the crystal and eventually lead to no crystal formation or bad crystal quality. Even though the crystal lattice force is typically weaker than the macromolecule folding energy, the overall arrangements of a protein with multi-domain or multimers may be affected by this weak lattice force (130,131). In these cases, SAXS can fill in those gaps, by providing the details of the overall conformation of the studied protein in solution, as well as with the flexible regions in the protein. By analyzing SAXS data of the full-length multi-domain protein or the multi-protein complexes, all the

available high-resolution domain/subunit structures can function as building blocks in the rigid-body modeling software, to obtain the overall conformation of the multi-domain protein and the quaternary structure of the multi-protein complex, and the missing regions can be rebuilt from the SAXS data.

NMR is another technique that provides atomic resolution structure but differs from X-ray crystallography in which NMR provides the structural information in solution. Aside from the structural information, NMR provides information about the dynamics and flexibility of the protein, as well as the interaction information between a protein-protein or protein-substrate. However, this technique requires expensive isotope labeling, and is limited by the molecular size (<70 kDa) that it can analyze, restraining the capability of this technique to study large proteins and protein assemblies. By complementing NMR with SAXS and X-ray crystallography, the structure, dynamics and interaction of the macro- or supramolecular complexes can be characterized in solution.

The capability of SAXS in handling a polydisperse system allows the study of the protein flexibility in solution. The SAXS data of a protein in solution is a sum of the intensities produced by each conformer. The structural study of the multi-domain protein, well-folded protein with disordered regions as well as intrinsically disordered protein can be achieved by considering all the conformers that can be attained from the SAXS data. The volume fractions of each component in the mixture can be retrieved from the collected SAXS data if the number of components are low. For a highly flexible system, where a lot of conformers are presented in the mixture, the ensemble approach such as Ensemble Optimization Method (EOM) are employed, to explain the system using an ensemble of conformations (132,133).

Apart from the static SAXS experiment, SAXS can be used to study the kinetics of the macromolecules, such as protein folding-unfolding transition that can occur from millisecond to hours and ligand-induced protein conformational change that can occur in microsecond to millisecond (134,135). With technological advancement in synchrotron, high beam intensity flux, fast detector readout and efficient mixing methods allow the investigation of conformational transition states, assembly processes and the kinetics of the protein in a time-dependent manner down to the time resolution of sub-millisecond (ms) (136-138). The continuous-flow mixer or the stopped-flow mixer provides a fast and homogenous mixing with a mixing dead time

less than 0.5 msec. Time-resolved SAXS (TR-SAXS) has been used to study the protein folding, RNA folding, aggregates formation, viral capsid self-assembly or viral RNA encapsidation (139-143). These experiments show that TR-SAXS can be a powerful tool to study the kinetics of the initial state and final state of the macromolecules in solution, and also to track down the intermediate states that may arise in the process.

1.13 Goals of this thesis

DENV NS5 is a protein with two functional enzymatic domains, tethered by a 10-residue linker. It plays a central role in DENV replication and to perturb the host immune response. The recently solved crystallographic structure of DENV NS5₆₋₈₉₅ (84) reveals a different domain-domain arrangement than the JEV NS5 structure (94). Analyzing the crystallographic packing of this crystallographic structure showed a tight crystal packing. To understand the domain-domain arrangement of the DENV full-length NS5 (NS5FL), as well as the potential dynamic in this two-domain protein, DENV-3 NS5FL was analyzed in solution using SAXS.

Depending on the antigenic properties of the viruses, DENV can be sub-grouped into four serotypes, named DENV-1, -2, -3 and -4. Previous experiments showed the differences in nuclear localization of the NS5 among four serotypes (73,74). To identify the structural similarities and differences, as well as the dynamic features, among the NS5FL of all four serotypes, these four proteins were investigated in solution using SAXS. Mutagenesis experiments were carried out with the NS5FLs to identify the critical residues causing the alterations observed.

In addition to the studies on substrate-free enzyme, the interaction between NS5FL and its substrates were analyzed in solution using SAXS, in order to track down any changes that may arise in the NS5FL upon substrate-binding or release. These interactions were further characterized using time-dependent SAXS experiments, to resolve the time scale that the change arises, as well as track down the possible transition states in NS5FL upon the binding or releasing of substrates.

The replication of DENV genome requires both NS5 and NS3. NS3 is also a protein with two functional domains connected via a flexible linker. The interaction between these two proteins are essential for DENV replication, and previous studies revealed that NS3-NS5 interaction is achieved through the helicase of NS3 and the RdRp of NS5. To get structural insight into the interaction between NS3 and NS5, the

DENV helicase (NS3₁₇₂₋₆₁₈) and a fused protein (NS3₁₇₂₋₆₁₈-NS5₃₂₀₋₃₄₁), containing the NS3 helicase and an NS5 peptide spanning residues 320 to 341, were studied in solution using SAXS. In addition, two domain-domain orientations were identified in the full-length NS3 from the DENV-4, as well as different orientations from the viruses coming from same family, suggesting that NS3 is flexible. To characterize this flexibility in solution, DENV-4 NS2B₁₈NS3 was analyzed using SAXS.

The structural protein, capsid (C) protein, is a dimeric protein that encapsidates the DENV RNA to form the nucleocapsid (33,34). The solution structure of this protein was solved using NMR technique, but due to the high flexibility at its N-terminus, the first 20 residues were not assigned into the reported structure (34). To gain an insight into this flexible region, the DENV-2 full-length C protein was studied using SAXS. Furthermore, the salt effect of this protein was also analyzed using SAXS to obtain a proper condition for capsid re-assembly.

In this thesis, the main structural technique used was SAXS. By using a home source SAXS machine, as well as the SAXS experiment conducted using synchrotron beam line, different aspects of the SAXS technique were utilized in characterizing DENV proteins, in order to extend the knowledge of information that this technique can provide. The home source SAXS machine allows the protein to be studied immediately after purification, and avoids the freezing-thawing process of the proteins which can induce stresses that may alter the protein trait(s) (144,145). The home source SAXS machine also allows the optimization of protein preparation for SAXS experiment prior to utilization of the expensive synchrotron facilities. However, due to the limitation of the X-ray flux from the home source, synchrotron radiation with high X-ray flux is required to acquire SAXS data at fast time-resolution with good signal-to-noise ratio, especially to time-dependent SAXS studies.

2. MATERIALS & METHODS

2.1 Materials

2.1.1 Chemicals

All the chemicals used for the study were of at least analytical grade. Chemicals were purchased from the following companies:

Buffers and salts	Sigma (St. Louis, MO, USA) USB (Sampscott, MA) Fluka (Sigma, Buchs Germany) Roth (Karlsruhe, Germany) Serva (Heidelberg, Germany)
DTT	Sigma (St. Louis, MO, USA)
Pefabloc ^{SC}	BIOMOL (Hamburg, Germany)
Agarose, SDS, glycine, APS etc.	Bio-Rad (Hercules, CA, USA)
Antibiotics	Calbiochem, Sigma, Gibco
IPTG	Gold Bio (St. Louis, MO, USA)

2.1.2 Molecular biology materials

Primers	AITbiotech (Singapore) IDT (Singapore)
<i>Pfu</i> DNA polymerase	Thermo Scientific (Waltham, MA, USA)
KAPA HIFI DNA polymerase	Kapa Biosystem (Massachusetts, USA)
<i>Bam</i> HI, <i>Nde</i> I, <i>Xho</i> I	Fermentas (Glen Burnie, MD, USA)
T4 DNA ligase	Promega (Wisconsin, USA)
PreScission TM Protease	GE Healthcare (Uppsala, Sweden)
QIAprep® Spin Miniprep Kit	QIAGEN (Hilden, Germany)
NucleoBond® AX Midiprep Kit	MACHEREY-NAGEL (Düren, Germany)
QIAquick® Gel Extraction Kit	QIAGEN (Hilden, Germany)
QIAquick® PCR Purification Kit	QIAGEN (Hilden, Germany)
<i>E. coli</i> plasmid propagation strain	DH5α
<i>E. coli</i> protein production strain	BL21(DE3) BL21-CodonPlus(DE3)-RIL
pET-24b	Novagen (Darmstadt, Germany)
pGEX6P-1	GE Healthcare (Uppsala, Sweden)

2.1.3 Chromatography

2.1.3.1 Affinity chromatography

Ni-NTA	QIAGEN (Hilden, Germany)
--------	--------------------------

2.1.3.2 Size exclusion chromatography

Superdex 75 10/300 GL	GE Healthcare (Uppsala, Sweden)
HiLoad 16/600 Superdex 75 pg	GE Healthcare (Uppsala, Sweden)
Superdex 200 10/300 GL	GE Healthcare (Uppsala, Sweden)
Superdex 200 Increase 10/300 GL	GE Healthcare (Uppsala, Sweden)

2.1.3.3 Instruments and accessories

ÄKTA FPLC	GE Healthcare (Uppsala, Sweden)
ÄKTA Pure	GE Healthcare (Uppsala, Sweden)
Minisart® Filters (0.45 µm)	Sartorius Stedim Biotech (Goettingen, Germany)
Syringe, needles and accessories	BD Biosciences

2.1.3.4 Protein concentration estimation

Amicon Ultra (10, 30, 50 kDa)	Milipore (Co-cork, Ireland)
Vivacon 500 (10, 50 kDa)	Sartorius Stedim Biotech (Goettingen, Germany)
BioSpec-nano micro-volume spectrophotometer	Shimadzu Corporation (Kyoto, Japan)

2.1.4 Other instruments

Biometra T personal PCR thermocycler	Biometra (Goettingen, Germany)
Biometra T gradient PCR thermocycler	Biometra (Goettingen, Germany)
Sonoplus HD2200 ultrasonicator	Bandelin (Berlin, Germany)
Micropulser electroporator	Bio-Rad (Hercules, CA, USA)
Ultraspec 2100Pro Spectrophotometer	GE Healthcare (Uppsala, Sweden)
MetalJet-NANOSTAR	Bruker AXS (Karlsruhe, Germany)

2.1.5 Computer software

Serial Cloner v2.6.1	© Franck Perez (SerialBasics)
PyMOL v1.5	Schrödinger, LLC (Portland, Oregon, USA)
OriginPro 9	OriginLab (Northampton, MA, USA)
SAXS v4.1.47	Bruker AXS (Karlsruhe, Germany)
ATSAS v2.7.1	EMBL Hamburg (Hamburg, Germany)
ScÅtter v2.3a	© Robert P. Rambo (Oxfordshire, UK)
SasTool	SSRL (California, USA)

2.2 Methods

2.2.1 Vector preparation (modified pET-24b)

The plasmid pET-24b (Novagen, Germany) was modified by replacing the T7-tag with an N-terminal His₆-tag. The conceptual map is shown in figure 2.2.1. The plasmid pET-24b was purified from DH5α *E. coli* cells carrying the pET-24b plasmid using the QIAprep® Spin Miniprep Kit (QIAGEN, Germany) as per the manufacturer's protocol. Isolated plasmid was digested with *NdeI* and *SacI* to remove T7-tag from the plasmid. Two DNA oligos were synthesized from Integrated DNA Technologies (IDT, Singapore) with sequences oligo1: 5'-TAT GCA CCA CCA CCA CCA CCA CGG ATC CAT GGG AAT TCG AGC T-3' and oligo2: 5'-CGA ATT CCC ATG GAT CCG TGG TGG TGG TGG TGC A-3'. These two oligos were firstly incubated at 95 °C for 5 min to relax the secondary structures and/or dimerization of DNA oligos, and then gradually cooled to room temperature for 1 hr. The annealed oligos were later subjected to phosphorylation at the 5'-end of the oligo by incubating them with T4 polynucleotide kinase (NEB, Massachusetts) at 37 °C for 30 min, followed by heat inactivation at 65 °C for 20 min. The phosphorylated oligos were then ligated to *NdeI*-*SacI* digested pET-24b, and transformed into DH5α *E. coli* cells. Due to the insertion of one *NcoI* restriction site into the multiple cloning sites of pET-24b, the modified pET-24b was extracted from DH5α *E. coli* cells and digested with *NcoI* and *SphI* restriction enzymes, to check the presence of the insert in the modified pET-24b vector. The extracted plasmid with the presence of the *NcoI*-*SphI* fragment was sequenced using the service provided by Axil Scientific (Singapore).

Figure 2.2.1: Plasmid pET-24b. (Left) Overview of plasmid pET24b from Novagen, Germany. (Right) Sequence comparison between the original pET-24b (top) and modified pET-24b (bottom). T7 tag was replaced with the His-tag, and *NcoI* restriction site was introduced into the modified plasmid.

2.2.2 Cloning of genes encoding viral proteins from DENV

2.2.2.1 Cloning of the genes of NS5 from DENV

The four plasmids encoding full-length non-structural protein NS5 from DENV-1 to DENV-4 were prepared by Dr. Yongqian Zhao (National University of Singapore, Singapore), and are listed in table 2.2.1.

	Strain	Plasmid	Residue no.	GenBank
DENV-1 NS5FL	Singapore	pProEx-HTb (Life technologies)	1 - 899	EU081230
DENV-2 NS5FL	New Guinea-C	pET15b (Novagen)	1 - 900	M29095.1
DENV-3 NS5FL	Singapore	pProEx-HTb (Life technologies)	1 - 900	AY662691
DENV-4 NS5FL	Singapore	pProEx-HTb (Life technologies)	1 - 900	GQ398256

Table 2.2.1 Construct information of full-length NS5 from DENV-1 to DENV-4

The constructs of single domains of NS5 (methyltransferase (MTase) and RNA-dependent RNA polymerase (RdRp)) generated are listed in table 2.2.2.

	With or without linker (residues 263-272)	Residue no.	GenBank
DENV-3 MTase ₁₋₂₇₂	Yes	1 - 272	AY662691
DENV-3 MTase ₁₋₂₆₂	No	1 - 262	AY662691
DENV-3 RdRp ₂₆₃₋₉₀₀	Yes	263-900	AY662691
DENV-3 RdRp ₂₇₃₋₉₀₀	No	273-900	AY662691
DENV-4 MTase ₁₋₂₇₂	Yes	1 - 272	GQ398256
DENV-4 MTase ₁₋₂₆₂	No	1 - 262	GQ398256

Table 2.2.2 Construct information of NS5 domains from DENV-3 and DENV-4

Genes encoding the constructs of single domains of NS5 were amplified using Polymerase Chain Reaction (PCR), with the primers (underlined: restriction sites) listed in table 2.2.3 and plasmids encoding DENV-3 NS5FL and DENV-4 NS5FL as the templates.

Constructs		Oligonucleotide sequence 5'-3'	T _m (°C)
DENV-3	Forward	ATG <u>GGA</u> <u>ATT</u> <u>CAC</u> GGA ACA GGC TCA	60.7
MTase ₁₋₂₇₂	Reverse	CCA <u>TAA</u> <u>GCT</u> <u>TCT</u> AGT TGG GTG TTT CTG G	59.1

DENV-3 MTase ₁₋₂₆₂	Forward	ATG <u>GGA ATT CAC</u> GGA ACA GGC TCA	60.7
	Reverse	CAT <u>GAA GCT TCT</u> ATC GAG TTC CTG CTC C	60.4
DENV-3 RdRp ₂₆₃₋₉₀₀	Forward	GAC <u>TGA ATT CAT</u> GTT AAT GCG GAA	53.9
	Reverse	CAC ATA <u>AGC TTC</u> TAC CAA ATG GCT CC	57.8
DENV-3 RdRp ₂₇₃₋₉₀₀	Forward	GAC <u>TGA ATT CTA</u> ATG GAT GTC ATT	51.0
	Reverse	CAC ATA <u>AGC TTC</u> TAC CAA ATG GCT CC	57.8
DENV-4 MTase ₁₋₂₇₂	Forward	CAT <u>GGG ATC CGG</u> AAC TGG GAC CAC AGG AG	66.6
	Reverse	CTG <u>AAA GCT TCT</u> AGT CTG GTT TTT CTG TTT CAG TGG	61.5
DENV-4 MTase ₁₋₂₆₂	Forward	CAT <u>GGG ATC CGG</u> AAC TGG GAC CAC AGG AG	66.6
	Reverse	CAT <u>GAA GCT TCT</u> ATC TCG TTC CTG CCC CA	63.1

Table 2.2.3 Primers used in PCR for the individual domains of NS5

PCR was set up on ice, in a total volume of 50 μ l, with the components shown in table 2.2.4.

Components	Amount (μ l)
Forward Primer (100 μ M)	0.5
Reverse Primer (100 μ M)	0.5
Plasmid DNA (50 ng/ μ l)	0.5
10x <i>Pfu</i> Buffer with 20 mM MgSO ₄	5.0
dNTP mix (10 mM)	1.0
Double-distillated water	41.5
<i>Pfu</i> DNA Polymerase (2.5 U/ μ l)	1.0

Table 2.2.4 Components of PCR reaction mixture

Pfu DNA polymerase was added as the last component into the PCR reaction mixture, and the mixtures were then placed into the PCR thermocycler using the PCR program shown in table 2.2.5.

Step	Temperature (°C)	Duration (sec)	
Initial denaturation	98	180	
Denaturation	98	30	} 30 cycles
Primer annealing	T _m – 5	45	
Primer extension	72	120/kb	
Final extension	72	300	

Table 2.2.5 PCR thermocycler program

After amplification, 5 μ l of PCR reaction mixture was applied onto an 1.5% agarose gel to check the presence and purity of the gene product. If the desired gene size was observed on the gel, gel extraction was performed by applying the remaining PCR

reaction mixture onto a preparatory agarose gel (1.5%). The correct amplified gene fragment was cut out and purified using the QIAquick Gel Extraction Kit.

The purified gene fragments were double-digested overnight at 37 °C, with restriction enzymes *EcoRI* and *HindIII* for DENV-3 MTase and RdRp, or *BamHI* and *HindIII* for DENV-4 MTase. The double-digested gene fragments were then purified using QIAquick PCR Purification Kit. Subsequently, the digested gene fragments were ligated into the modified pET-24b vector, and the ligation reaction was set up in a total volume of 10 µl, with the components stated in table 2.2.6.

Components	Amount
Vector (modified pET24b)	50 ng
Insert	Variable (5 times of vector)
T4 Ligase	1 µl
10X T4 ligase buffer	1 µl
Double-distilled water	Variable

Table 2.2.6 Components of ligation mixture

The ligation mixture was kept at room temperature for 3 hr, followed by butanol precipitation to desalt and concentrate the ligated products. 7 µl of sterile water was added to dissolve the precipitated DNA, which was subsequently used for the transformation into 80 µl of *E. coli* DH5a cells by electroporation.

Plasmid was isolated from the transformants using QIAprep Spin Miniprep Kit, and subsequently double-digested and applied onto an agarose gel (1.5%) to select the plasmid with correct insert size. The sequences of selected plasmids were confirmed by DNA sequencing (Axil Scientific, Singapore).

2.2.2.2 Cloning of genes encoding NS5 linker mutants from DENV-3 and DENV-4

The NS5 linker-swapped mutants were created by exchanging the linker residues from DENV-3 to DENV-4, and *vice versa*, as shown in figure 3.4.1A (Results section). The plasmids encoding DENV-3 NS5FL and DENV-4 NS5FL were used as templates. The primers encoding the mutations (underlined) are listed in table 2.2.7.

Constructs		Oligonucleotide sequence 5'-3'	T _m (°C)
D3NS5-d4linker	Oligo1	ACT CGA <u>AGT GTC TCC ACT GAA ACA GAA</u> <u>AAA CCA GAC</u> ATG GAT GTC ATT GGG GAA AGA	68.8
	Oligo2	TGT CAT <u>GTC TGG TTT TTC TGT TTC AGT</u> <u>GGA GAC ACT</u> TCG AGT TCC TGC TCC TAA ATC	68.0

D4NS5- d3linker	Oligo1	ACG AGA <u>CAT</u> GTT AAT GCG GAA CCA GAA <u>ACA CCC AAC</u> ATG ACA GTT ATT GGG AGA AGG	69.0
	Oligo2	TGT CAT <u>GTT</u> GGG TGT TTC TGG TTC CGC <u>ATT AAC ATG</u> TCT CGT TCC TGC CCC AAG	70.5

Table 2.2.7 Primers used in PCR for NS5 linker-swapped mutants

The NS5 linker mutants with amino acid substitution were created by substituting the single, double, triple and quadruple non-conserved linker residues from DENV-3 to DENV-4, as shown in table 3.4.2 (Results section). The plasmids encoding DENV-4 NS5FL were used as the templates. Constructs DENV-4 NS5 S264H, S266N/T267A and $\Delta S_{266}T_{267}$ were cloned by Ms. Ankita Pan, and the rest of the constructs were cloned by Dr. Ardina Grüber from our laboratory. The primers encoding the mutations (underlined) were listed as following:

Constructs		Oligonucleotide sequence 5'-3'	T _m (°C)
DENV-4 NS5 S264H	Oligo1	GGA ACG AGA <u>CAT</u> GTC TCC ACT GAA ACA	60.2
	Oligo2	GAA AAA CCA <u>AGT</u> GGA GAC ATG TCT CGT TCC TGC CCC AAG ATC CAC	68.4
DENV-4 NS5 K271T	Oligo1	ACG AGA AGT GTC TCC ACT GAA ACA GAA <u>ACA</u> CCA GAC ATG GAT GTC ATT GGG GAA AGA	69.3
	Oligo2	TGT CAT GTC TGG <u>TGT</u> TTC TGT TTC AGT GGA GAC ACT TCG AGT TCC TGC TCC TAA ATC	68.5
DENV-4 NS5 K271T/D273N	Oligo1	ACG AGA AGT GTC TCC ACT GAA ACA GAA <u>ACA</u> CCA <u>AAC</u> ATG GAT GTC ATT GGG GAA AGA	68.9
	Oligo2	TGT CAT <u>GTT</u> TGG <u>TGT</u> TTC TGT TTC AGT GGA GAC ACT TCG AGT TCC TGC TCC TAA ATC	68.1
DENV-4 NS5 T269N/K271T/ D273N	Oligo1	ACG AGA AGT GTC TCC ACT GAA <u>CCA</u> GAA <u>ACA</u> CCC <u>AAC</u> ATG ACA GTT ATT GGG AGA AGG	69.2
	Oligo2	TGT CAT <u>GTT</u> GGG <u>TGT</u> TTC <u>TGG</u> TTC AGT GGA GAC ACT TCT CGT TCC TGC CCC AAG	70.8
DENV-4 NS5 S264H/T269N/ K271T/D273N	Oligo1	ACG AGA <u>CAT</u> GTT TCC ACT GAA <u>CCA</u> GAA <u>ACA</u> CCC <u>AAC</u> ATG ACA GTT ATT GGG AGA AGG	68.9
	Oligo2	TGT CAT <u>GTT</u> GGG <u>TGT</u> TTC <u>TGG</u> TTC AGT GGA AAC <u>ATG</u> TCT CGT TCC TGC CCC AAG	70.5
DENV-4 NS5 S266N/T267A	Oligo1	AGA AGT GTC <u>AAC GCA</u> GAA ACA GAA AAA CCA GAC ATG	63.0
	Oligo2	TTC TGT TTC <u>TGC GTT</u> GAC ACT TCT CGT	67.4

		TCC TGC CCC	
DENV-4 NS5 $\Delta S_{266}T_{267}$	Oligo1	ACG AGA AGT GTC GAA ACA GAA AAA CCA	63.3
		GAC ATG ACA	
	Oligo2	TTC TGT TTC GAC ACT TCT CGT TCC TGC	65.8
		CCC AAG ATC	

Table 2.2.8 Primers used in PCR for DENV-4 NS5 linker mutants with amino acid substitution

PCR was set up on ice, in a total volume of 25 μ l, with the components shown in table 2.2.9. KAPA HiFi DNA polymerase was added as the last component into PCR reaction mixture, and the mixtures were placed into the PCR thermocycler and the PCR program was set as stated in table 2.2.10.

Components	Amount (μ l)
Oligo1 (100 μ M)	0.15
Oligo2 (100 μ M)	0.15
Plasmid DNA (10 ng/ μ l)	0.5
5x KAPA HiFi Buffer with 10 mM MgCl ₂	5.0
dNTP mix (10 mM)	0.75
Nuclease-free water	17.85
KAPA HiFi DNA Polymerase (1 U/ μ l)	0.5

Table 2.2.9 Components of PCR reaction mixture

Step	Temperature (°C)	Duration (sec)	
Initial Denaturation	98	120	
Denaturation	98	20	} 20 cycles
Primer Annealing	50, 55, 65, 68, 70	15	
Primer Extension	72	30/kb	
Final Extension	72	420	

Table 2.2.10 PCR thermocycler program

After amplification, 5 μ l of PCR reaction mixture was applied onto an agarose gel to check the presence and purity of the gene product. In order to remove the methylated template DNA, 10 μ l of PCR reaction mixture was incubated with DpnI at 37 °C for 1 hr (in a total volume of 50 μ l), followed by heat inactivation at 80 °C for 20 min. 5 μ l of the sample was then precipitated and transformed into DH5 α *E. coli* cell. Plasmids were isolated from the transformants using QIAprep Spin Miniprep Kit, and the sequences of mutated sites were confirmed by DNA sequencing (Axil Scientific, Singapore).

2.2.2.3 Cloning of gene encoding DENV-2 capsid protein

The viral cDNA encoding DENV-2 genomic RNA (Indonesia strain; GenBank: GQ398264), which was inserted into a BAC vector, was kindly provided by A/Prof. Liu Ding Xiang (School of Biological Sciences, Nanyang Technological University, Singapore). Gene encoding the construct DENV-2 capsid protein (residues 1-100) was amplified from the viral cDNA using forward primer 5'-GAT TGG ATC CAT GAA TAA CCA ACG GAA AAA GGC GAG AAA C-3' ($T_m = 63.9\text{ }^{\circ}\text{C}$) and reverse primer 5'-CAT GAA GCT TCT ATC TGC GTC TCC TGT TCA AGA TGT TCA G-3' ($T_m = 64.0\text{ }^{\circ}\text{C}$), with restriction sites BamHI and *HindIII* (underlined). The PCR reaction mixture was set up on ice with the components shown in table 2.2.4, and the PCR was done with the PCR program stated in table 2.2.5. The presence and purity of the gene product were verified by an agarose gel, and the gene was purified from the gel using the QIAquick Gel Extraction Kit. The purified gene product was double-digested with BamHI and *HindIII*, and ligated into the modified pET-24b vector. After butanol precipitation of the ligated product, the DNA was transformed into *E. coli* DH5 α cells by electroporation.

Plasmids were isolated from the transformants and double-digested with BamHI and *HindIII*, and applied to an agarose gel. The sequences of the plasmids with the correct insert size were then confirmed by DNA sequencing (Axil Scientific, Singapore).

2.2.2.4 Information of genes encoding DENV-4 NS3 protein and its domains

The pET32b plasmids encoding DENV-4 NS2B₁₈NS3 (GenBank: AY776330) (100) and NS3₁₇₂₋₆₁₈ (helicase) (115) were kindly provided by Asst. Prof. Dahai Luo (Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore). The construct NS2B₁₈NS3 consisting of the full-length NS3 (residues 1-618), and 18 residues from NS2B (residues 49-66) was physically linked to the NS3 via a Gly₄-Ser-Gly₄ linker. This NS2B peptide is required for proper folding in the protease domain of NS3. The pProEx-HTB vector encoding DENV-4 hel-NS5₃₂₀₋₃₄₁ was prepared by Dr. Moon Yue Feng Tay (Duke-NUS, Singapore) (91). This construct composed of the gene sequence of NS3 helicase domain (residues 172-618), and the NS5 peptide (320-341) was physically linked to the helicase via a Gly₄-Ser-Gly₄ linker.

2.2.3 Protein production of recombinant viral proteins from DENV

All the plasmids with the confirmed sequences were transformed into the protein production *E. coli* strain, BL21-CodonPlus(DE3)-RIL competent cells by electroporation at a constant voltage of 2.5 kV (Micropulser electroporator, Bio-Rad).

2.2.3.1 Induction test of viral proteins from DENV

Cells with genes encoding full-length NS5, NS5 mutants, DENV-4 NS2B₁₈NS3, helicase and hel-NS5₃₂₀₋₃₄₁ were grown in 4 ml Luria-Bertani (LB) media supplemented with 100 µg/ml ampicillin and 34 µg/ml chloramphenicol, at 37 °C with the shaking speed of 180 rpm. Cells with genes encoding MTase, RdRp and capsid protein were grown in 4 ml LB media supplemented with 30 µg/ml kanamycin and 34 µg/ml chloramphenicol, at 37 °C with the shaking speed of 180 rpm. When cell cultures reached an attenuation (OD₆₀₀) of 0.6, the cells were split into two aliquots, and the final concentration of 1 mM IPTG (Isopropyl-β-D-thiogalactoside) was added into one of them for protein production. The cells were then incubated at 37 °C for two hours, and subsequently centrifuged and re-suspended in 60 µl of 1x lysis buffer (0.25 M Tris-HCl, pH 6.8, 9.2% SDS, 40% glycerol, 0.2% bromophenolblue) in the presence of 12.5 mM DTT. The cell suspension was heated at 95 °C for 5 min, and 12 µl of sample was loaded onto a 17% SDS polyacrylamide gel (146) to check the protein production. After gel electrophoresis, the SDS polyacrylamide gel was stained with Coomassie staining solution (0.2% w/v Coomassie brilliant blue, 40% v/v methanol, 1% v/v acetic acid) for 5-10 min, and de-stained with de-staining solution 1 (25% v/v isopropanol, 10% v/v acetic acid), followed by de-staining solution 2 (7.5% v/v methanol, 10% acetic acid).

2.2.3.2 Protein production of viral proteins from DENV

Cells were grown in 1 liter LB media under the condition stated in section 2.2.3.1. Instead of performing protein induction at 37 °C, the protein was produced at 18 °C for 18 hours. After induction, the cells were harvested at 7,000 x g for 12 min at 4 °C.

2.2.4 Protein purification of viral proteins from DENV

To ensure high purity of proteins, all proteins were purified using a two-step purification protocol. After cell lysis, the His₆-tagged proteins were first separated from the cell lysate using Ni-NTA affinity chromatography, which is based on the concept of immobilized-metal affinity chromatography. In this step, nickel ions (Ni²⁺),

which are immobilized to agarose beads through nitrilotriacetic acid (NTA), bind His₆-tag that was engineered to the recombinant proteins, to separate them from *E. coli* proteins. After this, the protein purity was further polished using size exclusion chromatography (SEC), where the proteins were separated based on their molecular sizes. In this step, the protein oligomeric forms can be determined, and desired oligomeric states can be selected for downstream experiments.

2.2.4.1 Protein purification of NS5 constructs from DENV

The cells were lysed on ice by sonication (3 x 1 min) in lysis buffer (Table 2.2.11) with the addition of 2 mM Pefabloc^{SC} protease inhibitor.

	Lysis buffer	SEC buffer
NS5FL from all four serotypes	50 mM Tris/HCl, pH 7.5, 500 mM NaCl, 5% glycerol, 0.8 mM DTT	50 mM Tris/HCl, pH 7.5, 300 mM NaCl, 5% glycerol, 1 mM DTT
NS5FL linker mutants	50 mM Tris/HCl, pH 7.5, 500 mM NaCl, 5% glycerol, 0.8 mM DTT	50 mM Tris/HCl, pH 7.5, 300 mM NaCl, 5% glycerol, 1 mM DTT
MTase from DENV-3 and DENV-4	50 mM Tris/HCl, pH 7.5, 200 mM NaCl, 0.8 mM DTT	50 mM Tris/HCl, pH 7.5, 200 mM NaCl, 1 mM DTT
RdRp from DENV-3	50 mM Tris/HCl, pH 7.5, 500 mM NaCl, 5% glycerol, 0.8 mM DTT	50 mM Tris/HCl, pH 7.5, 200 mM NaCl, 1 mM DTT

Table 2.2.11 Buffer conditions used in protein purification

Cell debris were separated from supernatants by centrifugation at 12,500 x g for 25 min at 4 °C. The supernatant was filtered (0.45 µm) and the binding of proteins with Ni-NTA agarose was achieved using batch treatment. After incubation with Ni-NTA agarose for 60 min at 4 °C, the bound proteins were eluted from the Ni-NTA agarose beads with an imidazole gradient from 20 mM to 500 mM in the same buffer condition as the lysis buffer. The fractions containing the protein of interest was confirmed using SDS-PAGE, and were pooled together, concentrated, and then applied to a SEC column (Superdex 200 Increase 10/300 GL or Superdex 75 10/300 GL) with SEC buffer (Table 2.2.11).

2.2.4.2 Protein purification of NS3 constructs from DENV-4

The cells were lysed on ice by sonication (3 x 1 min) in buffer A (20 mM NaH₂PO₄/Na₂HPO₄, pH 7.5, 500 mM NaCl, 0.8 mM DTT), with addition of 2 mM Pefabloc^{SC}. Cell debris were separated from supernatants by centrifugation at 12,500 x

g for 25 min at 4 °C. The supernatant was filtered (0.45 µm) and the binding of proteins with Ni-NTA agarose was achieved using batch treatment. After incubation with Ni-NTA agarose for 60 min at 4 °C, the bound proteins were eluted from the Ni-NTA agarose beads with an imidazole gradient from 20 mM to 500 mM in buffer A. The fractions containing the respective proteins were confirmed using SDS-PAGE, and were pooled, and subjected to overnight protease cleavage. To cleave off the thioredoxin-His₆-tags, thrombin (Sigma-Aldrich, USA) was added into the protein samples containing NS2B₁₈NS3 or helicase in a ratio of 1 unit of thrombin to 500 µg of protein. To remove the His₆-tag, TEV protease was added into the protein sample containing hel-NS5₃₂₀₋₃₄₁ in a ratio of 1 unit of TEV protease to 50 mg of protein. The samples were dialyzed overnight at 4 °C against buffer B (20 mM NaH₂PO₄/Na₂HPO₄, pH 7.5, 200 mM NaCl). After overnight dialysis, the samples were concentrated, and applied onto a SEC column (Superdex 200 10/300 GL) in buffer C (20 mM Tris/HCl, pH 7.5, 200 mM NaCl, 1 mM DTT).

2.2.4.3 Protein purification of capsid protein from DENV-2

The cells were lysed on ice by sonication (3 x 1 min) in buffer D (50 mM Tris/HCl, pH 7.5, 1 M NaCl, 0.8 mM DTT) with the addition of 2 mM Pefabloc^{SC}. Cell debris were separated from supernatants by centrifugation at 12,500 x g for 25 min at 4 °C. The supernatant was filtered (0.45 µm) and the binding of proteins with Ni-NTA agarose was achieved using batch treatment. After incubation with Ni-NTA agarose for 60 min at 4 °C, the bound proteins were eluted from Ni-NTA agarose beads with an imidazole gradient from 20 mM to 1000 mM in buffer D. The fractions containing the recombinant protein was confirmed using SDS-PAGE, and were pooled together and applied directly onto a SEC column (HiLoad 16/600 Superdex 75 pg) with buffer D.

2.2.5 Small angle X-ray scattering (SAXS) experiments

SAXS experiments were performed by collecting the scattering data from a highly pure, monodisperse protein solution, as well as from the buffer solution. The buffer scattering is then removed from the protein scattering data. A few frames were collected for the same sample and compared to detect radiation damage in the sample that may arise during data collection. The protein scattering data were usually collected in a concentration range to identify potential concentration dependence (inter-particle interaction) of that protein. The potential aggregation was assessed

using Guinier plot. In general, for a protein sample that does not have inter-particle interaction, the scattering data from higher concentration, which has low noise, will be used for further data analysis. For a sample that has inter-particle interaction, the scattering data from different concentrations will be merged or extrapolated to an infinite concentration to obtain an interference-free scattering data that will be used for further data analysis.

2.2.5.1 SAXS data collection using home source NANOSTAR

SAXS data of DENV full-length NS5, NS5 mutants, MTase, RdRp, NS2B₁₈NS3, helicase, hel-NS5₃₂₀₋₃₄₁ and capsid protein were collected using a NANOSTAR SAXS instrument (Bruker AXS, Karlsruhe, Germany) equipped with MetalJet X-ray source (Excillum, Germany), as shown in figure 2.2.2. The X-ray radiation is generated after the impact of electron to a high-speed jet of liquid gallium alloy ($\lambda = 0.13414$ nm with a potential of 70 kV and a current of 2.857 mA), and collimated through Montel optics, followed by two 500 μm scatter-free pinhole system (Incoatec, Germany), to remove any parasitic scattering. A photon counting VÅNTEC-2000 detector (140 mm x 140 mm) (Bruker AXS, Germany) is used at a sample-to-detector distance of 0.67 m covering the range of momentum transfer $0.16 < q < 4 \text{ nm}^{-1}$, where $q = 4\pi \sin(\theta)/\lambda$, in which 2θ is the scattering angle and λ is the X-ray wavelength.

Figure 2.2.2: BRUKER NANOSTAR SAXS instrument located in the School of Biological Sciences, Nanyang Technological University. This instrument is equipped with a MetalJet X-ray source, a sample chamber with temperature control holder inside, and a VÅNTEC-2000 detector, and the beam is collimated with Montel optics and two scatter-free pinholes.

The scattering patterns of DENV recombinant proteins were collected at three protein concentrations. The samples were loaded into a vacuum tight quartz capillary and measured at 15 °C (288 K) temperature, with a sample volume of 30 μ l. For each sample, a total of six measurements in five minute intervals were collected. Using the built-in SAXS software (Bruker AXS, Germany), the data were flood-field corrected to produce a smooth detector data image by rectifying the non-uniformity in the pixel-to-pixel sensitivity, and spatial corrected to correct the inherent geometrical pincushion distortion. Then the 2D images were normalized with incident beam intensity, and integrated into an 1D-profile. Possible radiation damage was inspected by comparing all six data sets. If no radiation damage was detected, all six profiles were averaged to improve signal quality. Scattering of the buffer was subtracted from the protein scattering data, and normalized by protein concentration.

2.2.5.2 High-resolution SAXS data collection at Stanford Synchrotron Radiation Lightsource (SSRL)

The SAXS studies for DENV NS5FL in the presence or absence of substrates were carried out at the Bio-SAXS beamline BL4-2 of SSRL (147). The SAXS instrument at BL4-2 is shown in figure 2.2.3A.

Figure 2.2.3: SAXS instrument located in Stanford Synchrotron Radiation Lightsource (SSRL). (A) SAXS instrument at Beam Line 4-2 (BL4-2). (B) Close view of sample loading site. (C) The samples are loaded through the BL4-2 autosampler, a robotic automated sample delivery device.

High-resolution static SAXS data were collected using a Rayonix MX225-HE CCD detector (Rayonix, Illinois, USA) at a sample-to-detector distance of 1.7 m covering the range of momentum transfer $0.068 < q < 6.7 \text{ nm}^{-1}$, where $q = 4\pi \sin(\theta) / \lambda$, in which 2θ is the scattering angle and λ is the X-ray wavelength ($\lambda = 0.1127 \text{ nm}$), with a beam energy of 11 keV. A sample volume of 30 μl was loaded into a 1.5 mm quartz capillary cell using the BL4-2 autosampler (148) (Fig. 2.2.3C). Prior to the data collection, 5 mM of DTT, which acts as the radical scavenger, was added to all samples freshly, to protect the proteins from radiation damage. For each sample, a total of 15 frames with 1 s interval were recorded. All the data frames were integrated and buffer scattering intensity was subtracted from the sample data using the software SasTool (SSRL, California, USA). Possible radiation damage was inspected by comparing all 15 scattering profiles. The profiles without radiation damage were averaged for further data analysis.

2.2.5.3 Time-resolved SAXS (TR-SAXS) data collection at SSRL

Slow TR-SAXS was performed using the BL4-2 autosampler (Fig. 2.2.3C). The dead time of this setup was estimated to be 60 s. The TR-SAXS data were recorded using a Pilatus 300K single photon counting detector (200 Hz frame rate), with the beam energy same as static SAXS experiment (Section 2.2.5.2). After mixing the protein with the substrates in a total volume of 30 μl , the sample was injected using the autosampler. The first frame was acquired after 60 s, and a total of 15 frames with 1 s interval were recorded. Every 5 min, the data of the same sample was acquired, and data acquisition was stopped after 30 min. To reduce the radiation damage, the sample was wobbled inside the capillary. All the data frames were integrated and buffer scattering intensity was subtracted from the sample data using the software SasTool (SSRL, California, USA). Possible radiation damage was inspected by comparing all 15 scattering profiles of each time point and the profiles without radiation damage were averaged for further data analysis.

Fast time-resolved SAXS (TR-SAXS) was performed using a customized Bio-Logic SFM-400 four syringe stopped-flow mixer (Bio-Logic, Claix, France) (Fig. 2.2.4A), to achieve a time resolution as low as 5 ms. The mixing dead time of the stopped-flow mixer was estimated to be 0.25 ms. Higher photon flux was provided by the synthetic Mo/B₄C multilayer monochromator (149). TR-SAXS data were recorded using a Pilatus 300K single photon counting detector (200 Hz frame rate) (Dectris,

Baden, Switzerland). The samples containing protein and substrates were injected separately from two sample syringes (Fig. 2.2.4B), in a sample volume of 100 μ l per shot from each syringe. The sample injection from the stopped-flow mixer was controlled using Blu-ICE software (SSRL, California, USA). For each mixing event, the data for 50 images with 20 ms intervals (in total 1 s) were acquired. All the data frames were integrated and buffer scattering intensity was subtracted from the sample data using the software SasTool (SSRL, California, USA). Possible radiation damage was inspected by comparing all 50 scattering profiles.

Figure 2.2.4: Stopped-flow device located at SSRL. (A) Stopped-flow mixer (SFM-400) with four independent syringes and three mixers. (B) Schematic diagram of the customized stopped-flow mixer from SFM-400 at SSRL.

2.2.6 Small angle X-ray scattering (SAXS) data analysis

2.2.6.1 Primary SAXS data processing

The primary SAXS data processing steps were performed using the program package PRIMUS (150,151). To identify the presence of inter-particle interaction between protein molecules with increasing protein concentrations (concentration dependence), the scattering profiles from different protein concentrations were overlapped and compared at low angle region. If all scattering profiles overlapped nicely, it was evaluated as no concentration dependence in the scattering profiles. If the low angle of scattering profiles shifted upward or downward, it was evaluated that there is concentration dependence presented in the scattering profiles, and the data will either be extrapolated to infinite dilution or merged together.

Two important parameters, including forward intensity ($I(0)$) and radius of gyration (R_g), were determined from Guinier approximation and pair-distance distribution

($P(r)$) function. The forward intensity ($I(0)$) is the radiation intensity scattered through zero angle ($q = 0$). This parameter cannot be measured and it can only be determined through extrapolation, because the detected intensity cannot be differentiated from the un-scattered intensity that passes through the sample (130,152). $I(0)$ is not dependent on the particle shape and this value can be used to determine the molecular mass of a particle (130). The radius of gyration (R_g) is the root-mean-squared distance of each scatterer from the center of particle. This parameter is shape-dependent, whereby the objects with same volume but with different shapes have different R_g values (130,152).

The Guinier approximation (153,154) states that at very low angle ($q < 1/R_g$), the intensity can be represented as:

$$I(q) = I(0)e^{\frac{-q^2 R_g^2}{3}}$$

This means that when the scattering intensity are plotted as $\ln I(q)$ vs q^2 , the intensity exists as a linear function, and the R_g can be calculated from the slope of the curve and $I(0)$ can be derived from the y-intercept of the curve. However, this relationship is only valid for $qR_g < 1.3$ for globular protein. Protein aggregation and inter-particle interaction for all the protein samples were analyzed using the scattering data from the Guinier region (153), where the non-linear behavior at this region indicates the formation of aggregation. In time-resolved SAXS data analysis, this R_g value is computed using a command-line program AUTORG (155) to speed up the R_g estimation.

To validate these parameters, the R_g and $I(0)$ were also computed based on the pair-distance distribution ($P(r)$) function. This method considers the whole scattering data while computing these values and thus, the computation of both parameters can be more precise (152). The $P(r)$ function describes all the distances between two electrons in a protein particle as a paired-set, and can be computed using the software GNOM (156). This function is derived from the scattering intensity by an inverse Fourier transformation (156,157) and their relationship is represented as:

$$p(r) = \frac{r^2}{2\pi^2} \int q^2 I(q) \frac{\sin qr}{qr} dq$$

Additionally, the maximum dimension (D_{max}) of particle can also be determined from the $P(r)$ function, where D_{max} is the maximum pair-distance that can be found within the particle.

2.2.6.2 *Ab initio* modeling using SAXS

The *ab initio* low-resolution models of the proteins were reconstructed using two programs DAMMIF (158) and GASBOR (159). DAMMIF generates the *ab initio* models based on the bead modeling method, where the particle is represented as a collection of dummy atoms within a search volume, and each dummy atom can come from the particle or the solvent. A search volume is a sphere that has the diameter equal to the D_{max} determined from $P(r)$ function. This program considers the low angle data in a range of $q < 0.2 \text{ nm}^{-1}$. DAMMIF utilizes the simulated annealing to construct the bead model, and disconnected models are immediately discarded from the simulation, which speed up the simulation time. The calculation of theoretical scattering amplitudes of the models is only performed to the connected models, and a final model that has a lowest discrepancy (χ^2) between the theoretical scattering profile and the experimental data is selected. Second *ab initio* program, GASBOR is performed to verify the robustness of the collected data in shape analysis. This program is limited to samples with protein only, and it considers the full q -range in modeling. Instead of representing the protein particle by dummy atoms, GASBOR represents protein as an ensemble of dummy residues (DRs), where these dummy residues approximate the position of the α -carbon (C_α) atoms in the protein residues. Since the number of protein residues is known from the primary sequence, this number can serve as a constraint during the model reconstruction. Then the low-resolution model is reconstructed by searching a spatial arrangement of these DR chains within a spherical search volume of D_{max} . To generate a protein-like model, penalties are imposed to ensure that the model is connected and each DR has at least one DR at a distance of 0.38 nm. A final model that has the least χ^2 value between the simulated scattering profile and the experimental scattering profile was then generated.

To validate the reproducibility of the shape reconstructions, 20 independent *ab initio* reconstructions were performed for each scattering data, and all 20 reconstructed models were used as input for the program suite DAMAVER (160). DAMAVER has a set of programs that align these reconstructed models, select the typical model, and construct an averaged model. The normalized spatial discrepancy (NSD) values are computed between two models, where the two models are similar if the NSD is close to one. Subsequently, a final mean NSD over all model-pairs and their dispersion of NSD are calculated. The reconstructed models which have a high NSD dispersion are then discarded, while the rest of the models are averaged. In DAMMIF, the averaged

ab initio model was selected as the representative model, while in GASBOR, the typical *ab initio* model, which has the least NSD value, was selected as the representative model. Beside this, the averaged excluded volume (V_{Ex}) computed from DAMAVER using the DAMMIF models as input, was used to estimate the molecular mass of the protein (161).

2.2.6.3 Protein folding and flexibility assessment using SAXS

To qualitatively access the globular nature of the particle, Kratky plot can be generated by transforming the scattering pattern of the particle as $q^2 I(q)$ vs q . For a globular protein with a well-defined and solvent-accessible surface, its Kratky plot will exhibit a typical bell-curve profile with a well-defined maximum. Conversely, for a random chain, Kratky plot exhibits a plateau or slightly increase at high q region. The Kratky plot cannot distinguish clearly between partially unfolded and fully folded particles since both exhibit similar profile. To overcome this, a normalized Kratky plot is generated ($(qR_g)^2(I(q)/I(0))$ vs qR_g), by normalizing the intensity $I(q)$ with the forward intensity $I(0)$, and q with the radius of gyration (R_g) of the particle (130,162-164). Therefore, the normalized Kratky plot is independent to the particle size and molecular mass, where $I(0)$ is related to the particle molecular mass. This allows one to compare the well-folded globular particle, extended folded particle and fully disordered particle. For the well-folded globular particle, despite their particle sizes, their normalized Kratky plots exhibit a well-defined bell-curve shape, with a maximum $(qR_g)^2(I(q)/I(0))$ value of 1.104 at qR_g value of $\sqrt{3}$. For the fully disordered particles, their normalized Kratky plots lose the bell-curve shape, and the $(qR_g)^2(I(q)/I(0))$ values increase with the increased qR_g values. For the particles exhibiting the shape in between, their normalized Kratky plots exhibit a profile that lies in between these two extremes.

By transforming the scattering data as $q^4 \cdot I(q)$ vs q^4 , a Porod-Debye plot can be generated, and it approximates the scattering intensity for a well-folded compact particle decay as q^{-4} scaled by the particle's surface area, concentration and electron density contrast within a limited angle range (165). The compact particle produces a sharp scattering contrast (the electron density contrast between the particle and solvent) and subsequently, displays a plateau in the Porod-Debye plot. This plateau will dissipate with an increasing flexibility in particle due to the diffusion in scattering contrast.

2.2.6.4 Rigid-body modeling using SAXS

Rigid-body modeling was performed for the NS5 of the four DENV serotypes using the software BUNCH (166), by docking the high resolution crystallographic structures of single domains (PDB IDs: 3P8Z (MTase) (79) and 4C11 (RdRp) (86)) against the experimental data. BUNCH performs the modeling using a combination of rigid-body and *ab initio* modeling approach. The high-resolution structures act as the rigid bodies, and BUNCH employs a simulated annealing procedure to search for the optimal orientations of these rigid bodies, and connects them with the dummy residue (DR) chains that correspond to the flexible linkers, to simultaneously fit the experimental scattering data. Then the theoretical scattering intensity from the high resolution domain structures, together with its DR-chain, are computed using CRY SOL (167) and compared with the experimental scattering intensity, generating the discrepancy χ^2 value. The χ^2 value is used to analyze the goodness of fit between the experimental and calculated data, and is defined as:

$$\chi^2 = \frac{1}{N_k - 1} \sum_{j=1}^{N_k} \left[\frac{I(q_j) - cI_{calc}(q_j)}{\sigma(q_j)} \right]^2$$

N_k is the number of experimental point, c is the scaling factor, $I(q_j)$ is the intensity of experimental data at point j , and $I_{calc}(q_j)$ is the intensity of calculated data at point j . A better fit between the experimental and calculated data gives rise to a lower χ value. Therefore, in most of the data fitting and model building programs, χ value is minimized to generate a best solution that explains the experimental data. The rigid-body models are superimposed into the reconstructed low resolution envelopes using the software SUPCOMB (168), where it aligns two models by minimizing the normalized spatial discrepancy (NSD), which is a qualitative measurement of the similarity between the sets of three-dimensional points from two models.

Rigid-body modeling of the DENV-2 capsid protein was performed using the program CORAL (161). CORAL combines approaches utilized in both SASREF (166) and BUNCH. Similar to SASREF, CORAL allows the translation and rotation of the domains from multiple components of a complex. However, unlike SASREF, which explores the conformational space by random rotation of domains and does not account for the missing residues, the N- and the C-terminus of the domains are constrained with a pre-generated library that consists of the self-avoiding DR random loops (length from 5 to 100 residues) in CORAL. If no linker with appropriate length

to connect two domains is found, this conformation is rejected. For successful queries, the random loops are inserted into the domains, and its scattering intensity will be computed to compare with the experimental data (161).

2.2.6.5 Ensemble modeling using SAXS

For a system containing the homogenous random particles in solution, the observed scattering pattern is the sum of the scattering patterns produced by each particle averaged over all the orientations. Meanwhile, for a system containing flexible particles that adopts different conformations in solution, each conformation will contribute significantly to the final scattering pattern. Therefore, the SAXS data of NS5 were further analyzed by considering an ensemble of conformations. The Ensemble Optimization Method (EOM) (132) quantitatively assesses the flexibility that is present in the particle, by fitting the averaged theoretical scattering pattern from an ensemble of conformations to the experimental scattering pattern. Using the updated EOM version, EOM 2.0 (133), a random pool of 10,000 independent models was generated in order to approximate the conformational space. In this step, the crystallographic structures of the MTase and the RdRp were used as the inputs, and the linker region and missing loops in the high resolution structures were created based on the given protein sequences. For each model presented in the random pool, the theoretical scattering pattern is automatically computed using CRY SOL. After that, genetic algorithm (GA) is performed to select the subsets of conformations from the random pool and compared with the experimental scattering data, to formulate the discrepancy (χ^2) between two scattering patterns. The final ensemble with the minimum discrepancy was then selected. For each protein, the GA was repeated 100-times, and the averaged R_g and D_{max} from each final ensemble were computed and represented as R_g and D_{max} distributions. The final ensemble with the least discrepancy was selected and the conformations were reported as representative models. The distribution of R_g values describes the system flexibility. The broad R_g distribution of selected pool indicates the randomness of the particle, while the narrow R_g distribution suggests that the particle is rigid. Besides this, the distributions with R_g average values above the R_g average values from the random pool are considered as extended, whereas models with values below the average values as compact. When this is applied to the multi-domain proteins, the former indicates a long inter-domain distance, where the domains are away from each other, and the latter indicates a short

inter-domain distance, where the domains are closer to each other. This qualitative characterization of the particles can be quantified as R_{flex} , where the selected ensemble R_{flex} values close to the one extracted from the random pool R_{flex} are considered as an indication of random motion, while the R_{flex} values smaller or larger than the one extracted from the random pool R_{flex} indicate a rigid or flexible system. The complementary metrics R_{σ} represents the ratio of the standard deviations for the selected pool and random pool. The R_{σ} value is expected to be less than one when the selected ensemble R_{flex} is smaller than the random pool R_{flex} .

2.2.7 Protein sequence analysis

Multiple sequence alignments (MSAs) were performed using an interactive sequence alignment viewing, editing and analysis tool, Jalview 2.9 (169). The protein sequences were retrieved from the Virus Pathogen Database and Analysis Resource (ViPR) (170) and the UniProt database (171). These sequences were submitted to the web-server, Clustal Omega ((172); <http://www.clustal.org/omega/>) via Jalview 2.9, to perform MSA. The MSA results were analyzed in Jalview 2.9, and the sequence similarity was colored according to the percentage of the residues in each column that agree with the consensus sequence.

3. RESULTS

3.1 Sequence analysis of NS5

3.1.1 Sequence analysis of NS5 from DENV-1-4

The protein sequences of full-length non-structural protein 5 (NS5) from DENV-1 (GenBank: EU081230), DENV-2 (GenBank: M29095.1), DENV-3 (GenBank: AY662691) and DENV-4 (GenBank: GQ398256), were retrieved from the Virus Pathogen Database and Analysis Resource (ViPR) (170). To analyze the sequence conservation of the NS5 across different serotypes, multiple sequence alignments were performed using Clustal Omega (172) from Jalview 2.9 (169), with the retrieved protein sequences as input. The alignment outcome is shown in figure 3.1.1, and the secondary structure derived from the crystallographic structure of DENV-3 NS5₆₋₈₉₅ (PDB ID: 4V0R (84)) is included on top of the sequence alignment. The alignments show that both NS5 domains, MTase (*blue*) and RdRp (*magenta*) are highly conserved among all four serotypes, which have a sequence similarity of 74.1-83.2% and 75.4-81.5%, respectively. Despite the presence of high sequence conservation in both domains, the sequence similarity of the 10-residues linker region (*red dashed box*), which connects the two domains, is only 30-60%.

In the MTase domain, two protein stretches, which are important for MTase function, are highlighted in the figure 3.1.1 as *blue dashed boxes*. The AdoMet (SAM; S-Adenosyl methionine) binding site is almost fully conserved among all four DENV serotypes. While in the GTP-binding site, the important residues responsible for GTP-binding (L17, N18, L20 and F25) (50) are conserved among all DENV serotypes. The catalytic tetrad required for 2'-O-MTase activity, K61, D146, K181 and E217 (*red arrows* in Fig. 3.1.1) (50) are conserved among DENV-1 to DENV-4. NS5 is proposed to interact with NS3 (69) via the MTase residues 187-202 (*blue dashed box* in Fig. 3.1.1) (97). Across all four serotypes, the sequences in this interacting site are also highly conserved (66).

In the RdRp domain, the interacting sites for both importin β and NS3 helicase, are highly conserved among all four serotypes (*magenta box* in Fig. 3.1.1, labeled as bNLS/NS3 helicase interacting site) (71,90,91), while lower sequence similarity was observed for the a/bNLS sequences (70,71). This dissimilarity was discussed previously (73,74), where NS5s from different serotypes were shown to be localized in different cell compartments. These studies showed that the NS5s from DENV-1 and DENV-4 preferably localize in the cell cytoplasm, and DENV-2 and DENV-3 localize in the cell nucleus. This difference may be caused by the non-functional a/bNLS in

DENV-4 (73,74). The priming loop (highlighted as *magenta box* in Fig. 3.1.1) that is required for the stabilization of the RNA initiation complex is highly conserved among the serotypes.

Figure 3.1.1 Multiple sequence alignments of full-length NS5 from all four serotypes, adapted from (173). NS5 from DENV-1 (GenBank: EU081230), DENV-2 (GenBank: M29095.1), DENV-3 (GenBank: AY662691) and DENV-4 (GenBank: GQ398256) were aligned using Clustal Omega (172). The secondary structures were derived from the DENV-3 NS5₆₋₈₉₅ crystallographic structure (PDB ID: 4V0R (84)), and labeled as boxes for α -helices and arrows for β -strands. The MTase and RdRp were colored in blue (dark blue, N-terminal GTP-binding subdomain; blue, core subdomain; light blue, C-terminal subdomain) and magenta (magenta, finger subdomain; dark magenta, palm subdomain; light magenta, thumb subdomain), respectively, and 10-residue linker as red dashed box. The four catalytic residues responsible for 2'-O-MTase are indicated as red arrows below the sequences. The positions of RdRp conserved motifs are indicated as magenta line below the sequences (174). The protein stretches discussed in the text are labeled as dashed box.

3.1.2 Sequence analysis of NS5 from flaviviruses

The DENV NS5 amino acid sequences were further compared with NS5 from the same genus viruses (Flavivirus), including yellow fever virus (YFV; GenBank: X03700), West-Nile virus (WNV; GenBank: M12294), Kunjin virus (GenBank: AY274504), Japanese encephalitis virus (JEV; GenBank: M55506) and Zika virus (GenBank: KJ776791), and their protein sequences were retrieved from the UniProt database (171). The multiple sequence alignment was performed using the same server as stated in section 3.1.1. As shown in figure 3.1.2, a few conserved protein stretches were identified within the MTase (*blue*) and the RdRp (*magenta*), while very low sequence similarity was seen in the 10-residue linker region (*red dashed box*).

In the MTase, the AdoMet-binding site appears to have the best sequence conservation in NS5 across all different flaviviruses. The catalytic tetrad (K61, D146, K181 and E217) (*red arrows* in Fig. 3.1.2), which is important for the 2'-O-MTase activity is also conserved among all selected flaviviruses. Despite low sequence similarity in the GTP-binding site, the residues responsible for GTP-binding, L17, N18 and F25 are fully conserved among all selected flaviviruses, while L20 is conserved among DENV NS5 and yellow fever virus (YFV). However, this residue is replaced by a methionine in NS5 of West-Nile virus (WNV), Kunjin virus, Japanese encephalitis virus (JEV) and Zika virus. For the interacting site within the NS3 protease, the protein sequences are not fully conserved among the selected flaviviruses, but the charged residues that may play an important role in the interaction, are conserved.

In the RdRp, six highly conserved motifs that are important for RdRp catalytic activity (binding of metal ions, nucleotides and RNA, and phosphoryl transfer), are highlighted in figure 3.1.2. Similar to the observation seen in the sequence alignment shown in section 3.1.1, the interacting site for importin- β and NS3 helicase is highly conserved among the selected flaviviruses, while the α /bNLS-binding site shows low sequence conservation. Comparing the amino acid sequences of the priming loop, the important residue W795, which is involved in stabilizing the interaction between priming nucleotides and RNA template (50), is conserved among all selected flaviviruses.

Figure 3.1.2 Multiple sequence alignments of NS5 from different flaviviruses. NS5 from DENV-1 (GenBank: EU081230), DENV-2 (GenBank: M29095.1), DENV-3 (GenBank: AY662691), DENV-4 (GenBank: GQ398256), yellow fever virus (YFV; GenBank: X03700), West-Nile virus (WNV; GenBank: M12294), Kunjin virus (GenBank: AY274504), Japanese encephalitis virus (JEV; GenBank: M55506) and Zika virus (GenBank: KJ776791) were aligned using Clustal Omega (172). The secondary structures were derived from the DENV-3 NS5₆₋₈₉₅ crystallographic structure (PDB ID: 4V0R (84)), and labeled as boxes for α -helices and arrows for β -strands. The MTase and RdRp were colored in *blue* (*dark blue*, N-terminal GTP-binding subdomain; *blue*, core subdomain; *light blue*, C-terminal subdomain) and *magenta* (*magenta*, finger subdomain; *dark magenta*, palm subdomain; *light magenta*, thumb subdomain), respectively, and the 10-residue linker as *red dashed box*. The four catalytic residues responsible for the 2'-O-MTase are indicated as *red arrows* below the sequences. The positions of RdRp conserved motifs are indicated as *magenta line* below sequences (174). The protein stretches discussed in the text are labeled as *dashed box*. The critical residue in the priming loop is labeled as *green arrow* below the sequences.

3.2 Characterization of full-length NS5 (NS5FL) from DENV-3 in solution

NS5 (residues 1-900) from dengue virus (DENV) is a protein with two domains that catalyzes two different enzymatic reactions. The MTase is responsible for the capping of virus genome, while the RdRp is required for virus genome replication. These domains are connected via a 10-residue linker. The crystallographic structure of DENV-3 NS5₆₋₈₉₅ (PDB ID: 4V0R) showed that the MTase is arranged above the RdRp finger subdomain (84), as shown in figure 3.2.1A. The 4V0R structure has a temperature factor (B-factor) distribution as shown in figure 3.2.1B, where the thumb subdomain of RdRp has a high B-factor compared to the rest of the structure, indicating that this region has higher mobility in the protein crystal. However, the 10-residue linker region (*black arrow* in Fig. 3.2.1B) has a low B-factor, showing low mobility of this region in the crystallographic structure.

Figure 3.2.1 Crystallographic structure of DENV-3 NS5₆₋₈₉₅ (PDB ID: 4V0R) (84). (A) Overall structure of NS5₆₋₈₉₅ in *cartoon representation*. The MTase colored in *blue*, the 10-residue linker (residues 263 to 272) in *black* and the RdRp is colored according to their subdomains, where the finger subdomain is *red*, the palm subdomain is *green* and the thumb subdomain is *violet*. The proposed protein stretch interacting with NS3 is colored in *yellow*. Two substrates, GTP (Guanosine triphosphate) and SAH (*S*-adenosyl-homocysteine) in MTase are represented as *sticks* in *green*. The overall dimensions of 6.8 x 7.1 x 5.8 nm are estimated using PyMOL. (B) The putty representation of NS5₆₋₈₉₅ colored by B factor from low (blue) to high (red). The linker region is pointed by the *black arrow*. This figure is taken from Saw *et al.*, 2015 (173).

Looking into the crystal packing of the NS5 structure, it reveals a tight crystal packing, as shown in figure 3.2.2. This tight packing may reduce the freedom for the

domain-domain movement, and this may partially explain the low B-factor in the linker region of the crystallographic structure. In order to get further insights into the arrangement of two NS5 domains and track down the potential dynamics of NS5 in solution, DENV-3 NS5FL was analyzed using small-angle X-ray scattering (SAXS), together with its single domains.

Figure 3.2.2 Crystal packing of 4V0R structure of DENV-3 NS5₆₋₈₉₅ (84). NS5₆₋₈₉₅ is shown in *cartoon representation*. The MTase colored in *blue*, the 10-residue linker (residues 263 to 272) in *black* and the RdRp in *magenta*. NS5₆₋₈₉₅ in the 4V0R structure has a tight packing.

3.2.1 Expression and purification of DENV-3 NS5FL

Induction test was performed for DENV-3 NS5FL to ensure the protein is produced in *E. coli* cells, which was confirmed by a SDS gel. DENV-3 NS5FL was purified as described in section 2.2.4.1 (Materials and Methods). Recombinant NS5FL eluted from the Ni-NTA column at an imidazole concentration of 50 mM to 150 mM. Impurities, which may probably represent degraded products, were co-purified in this step, as shown in the SDS gel of figure 3.2.3A. In order to remove these impurities, and possible oligomeric forms of NS5FL, samples were concentrated and applied to size exclusion chromatography (SEC). Four peaks were identified from the SEC

elution profile (Fig. 3.2.3B). Except peak II, which had low absorbance, the fractions from peak I, III and IV were applied to a SDS gel (*inset* of Fig. 3.2.3B). The impurities were identified at peak I and IV, and DENV-3 NS5FL was identified to be at peak III. The SDS gel shows high purity of the protein. The elution fractions corresponding to the top 15% of peak III (*grey shaded area* in Fig. 3.2.3B) were pooled and concentrated to three protein concentrations for small-angle X-ray scattering (SAXS) experiments.

Figure 3.2.3 Purification of DENV-3 NS5FL. (A) SDS gel after Ni-NTA affinity chromatography. The samples from 50-150 mM imidazole fractions were pooled and concentrated, and applied to a SEC column to achieve higher purity sample. (B) The SEC elution profile of DENV-3 NS5FL. Four peaks were identified in the elution profile, and the presence of DENV-3 NS5FL was identified at peak III. (*inset*) SDS gel loaded with the fractions from peak I, III and IV. The impurities were identified at the fractions from peak I and IV. Peak III fraction contained high purity of DENV-3 NS5FL.

3.2.2 Primary SAXS data analysis of DENV-3 NS5FL

To ensure that the protein quality would not be affected by freezing and thawing, the SAXS data of DENV-3 NS5FL were collected immediately after the last purification step. SAXS data of NS5FL were recorded at 1.0, 2.1 and 4.1 mg/ml concentration. The SAXS curves at low angle shifted upward with increasing concentrations (Fig. 3.2.4B), indicating the presence of inter-particle interaction between NS5FL proteins. To remove this concentration-dependent effect, the data were extrapolated to theoretical infinite dilution and used for data analysis (Fig. 3.2.4A). The overall structural parameters derived from these SAXS experiments are shown in table 3.3.1. To assess the data quality of NS5FL samples, Guinier plots of all three concentrations at very low angles were prepared and all of them showed linearity, indicating good data quality for all the samples and no aggregation were seen in all

datasets (*top inset* of Fig. 3.2.4A). The Kratky plots of DENV-3 NS5FL showed a bell-curve profile with a well-defined maximum, indicating that the protein was well-folded (*bottom inset* of Fig. 3.2.4A). Based on Guinier approximation, the R_g value of DENV-3 NS5FL was determined to be 3.37 ± 0.02 nm. In comparison, the R_g value computed from the pair-distance distribution ($P(r)$) function, which considers the data from the whole scattering curve, was determined to be 3.47 ± 0.02 nm, confirming the R_g value derived from Guinier approximation. The $P(r)$ function of DENV-3 NS5FL shown in figure 3.2.4C had a bell-curve shape with a small tail from 8 nm till 11.5 nm, a trend that is usually seen for the protein with a slightly elongated conformation. The maximum particle dimension D_{max} of DENV-3 NS5FL protein was determined to be 11.5 ± 1 nm.

Figure 3.2.4 Primary SAXS data analysis of DENV-3 NS5FL. (A) SAXS scattering pattern (*black circle*) of DENV-3 NS5FL at theoretical infinite dilution. The Guinier plots (*top inset*) of all protein concentrations showed linearity, indicating no protein aggregation. The Kratky plot (*bottom inset*) showed that NS5FL was well-folded. (B) Comparison of SAXS scattering patterns at very low angle for all three measured protein concentrations. The upward shift with increasing concentration indicated the presence of inter-particle interaction between NS5FL. (C) Pair-distance distribution ($P(r)$) function of DENV-3 NS5FL.

3.2.3 Reconstruction of *ab initio* models of DENV-3 NS5FL

The shape of NS5FL of DENV-3 was reconstructed *ab initio* from the scattering pattern shown in figure 3.2.4B using the programs DAMMIF (158) and GASBOR (159). Twenty independent reconstructed DAMMIF models were computed to validate the reproducibility and robustness of the data and the algorithm used to compute

models. To validate these, the normalized spatial discrepancy (NSD), which is a quantitative estimation of similarity between sets of three-dimensional points of two models (168), was computed for all 20 models using the program DAMAVER. The NSD will be less than or close to 1 for two similar objects, while a high NSD value is expected for two systematically different objects. For a bead modeling software like DAMMIF, the computed NSD is in general smaller than or equal to one. The NSD of 20 NS5FL models were calculated to be 0.810 ± 0.053 . The $NSD < 1$ and low standard deviation indicated that the shape reconstructions of NS5FL protein were reproducible and hence it validated the robustness of the DAMMIF calculation using the collected dataset. The corresponding fit of the DAMMIF model and experimental data for NS5FL had a discrepancy (χ^2) of 0.672. The averaged model of DENV-3 NS5FL is shown in figure 3.2.5A, revealing an elongated shape, with a slight bend in the middle of the shape.

Figure 3.2.5 *Ab initio* modeling of DENV-3 NS5FL. (A) The averaged DAMMIF envelope of DENV-3 NS5FL from 20 independent reconstructions, with front (*left*) and side (*right*) views. The overall envelope dimensions were 10.5 x 7 x 4 nm. (B) The GASBOR envelope of DENV-3 NS5FL with best NSD value from 10 independent reconstructions, with front (*left*) and side (*right*) views.

To verify the DAMMIF envelope, a second *ab initio* program, GASBOR, was used for shape analysis. Compared to DAMMIF, which uses the scattering data up to $q=2 \text{ nm}^{-1}$, GASBOR utilizes the data from the whole scattering pattern. Ten independent GASBOR calculations yielded the NSD value of 1.276 ± 0.050 . The corresponding fit of the GASBOR model and experimental data for NS5FL resulted in a χ^2 of 0.580. The GASBOR model with the least NSD was shown in figure 3.2.5B, which corresponded to the DAMMIF envelope, eliciting an elongated shape with a slight

bending in the middle and a calculated NSD value of 0.873. Similar overall shapes were obtained from both methods, showing that the shape determination is reliable.

3.2.4 Structural comparison between the SAXS envelope and crystallographic structure

Both the $P(r)$ function and *ab initio* reconstruction of DENV-3 NS5FL showed that the protein has an elongated structure in solution. To get further insight into the overall structure of this two-domain protein, the crystallographic structure of DENV-3 NS5₆₋₈₉₅ (PDB ID: 4V0R (84)) was retrieved from the RCSB PDB (129). Investigating the arrangement of the two domains in the crystallographic structure revealed a more compact conformation of DENV-3 NS5₆₋₈₉₅, with overall dimensions of 8.7 x 7.2 x 5.5 nm. Using both the crystallographic structure and the solution scattering data as input into the program CRY SOL, the fitting between the experimental data and the theoretical scattering data calculated from the crystallographic structure resulted in a χ^2 value of 3.856 (Fig. 3.2.6A), and low agreement at the low angle region ($q < 0.35 \text{ nm}^{-1}$) and the angle range of 0.75 to 1.25 nm^{-1} (red arrow in Fig. 3.2.6A).

Figure 3.2.6 Rigid body modeling of DENV-3 NS5FL. (A) SAXS scattering pattern (black circle) of DENV-3 NS5FL and calculated scattering profiles of the crystallographic structure 4V0R (84) (red line) and the BUNCH model (blue line). The red arrows indicated the regions that showed deviation in the 4V0R calculated scattering profile (red line) with the experimental scattering pattern (black circle). (B) Superimposition of the DAMMIF envelope (grey sphere) and the BUNCH model (cartoon representation). The MTase and RdRp were represented as cartoon, and colored in blue and magenta, respectively. The missing residues, which were re-built, were represented as grey spheres.

Considering the differences seen between the experimental data and the crystallographic data, computational docking of the individual domains, MTase and RdRp, against the experimental SAXS data was done using the rigid-body modeling program, BUNCH (166). Similar to *ab initio* modeling, ten independent calculations

were performed to assess the reproducibility of the rigid-body model. All ten BUNCH models showed a similar arrangement of the MTase with respect to the RdRp, with an NSD of 1.324 and a low standard deviation of 0.035. The BUNCH model with the least χ^2 of 0.85 was used as a represented model, which was superimposed into the DAMMIF envelope using SUPCOMB (Fig. 3.2.6B). Instead of positioning the MTase on top of the finger subdomain of the RdRp as seen in the crystallographic structure (84), the MTase in the BUNCH model was positioned next to the RdRp to form a more elongated structure. The lower χ^2 value in the BUNCH model showed a good agreement with the experimental data (Fig. 3.2.6A), indicating that an elongated conformation is preferable in solution.

3.2.5 Domain mobility of DENV-3 NS5FL

The significant difference in the arrangement of the MTase and the RdRp in solution compared to the NS5₆₋₈₉₅ crystallographic structure (PDB ID: 4V0R) led to the consideration of domain mobility in DENV-3 NS5FL. To extract this information, normalized Kratky plot ($(q/R_g)^2 I(q)/I(0)$ vs q/R_g) was computed, where the plot is independent from the particle size and particle concentration. The normalized Kratky plot (or dimensionless Kratky plot) was plotted as shown in figure 3.2.7A using the scattering pattern of DENV-3 NS5FL, in comparison with the normalized Kratky plot of the standard globular protein, lysozyme. The normalized Kratky plot of lysozyme had a bell-shape profile with a maximum at $(\sqrt{3}, 1.104)$, indicating a properly folded globular protein. The maximum of the bell curve in the normalized Kratky plot of DENV-3 NS5FL shifted slightly right and upwards, showing a feature that was commonly seen in multi-domain proteins tethered by a linker with a slightly extended conformation (162,175), which allows some flexibility to the protein.

To assess the potential flexibility in this protein, the scattering data of DENV-3 NS5FL was used to plot the Porod-Debye plot ($q^4 I(q)$ vs q^4). In addition, the Porod exponent was calculated using the program ScÅtter. Following the Porod-Debye law, a compact particle will give rise to a sharp scattering contrast, which in turn shows a plateau in the Porod-Debye plot, while for a particle with increasing flexibility, the scattering contrast will diffuse, causing the loss of the Porod plateau (165). As shown in figure 3.2.7B, the Porod-Debye plot of DENV-3 NS5FL had an increasing slope at the Porod-Debye region, with a Porod exponent of 3.7, which was slightly lower

compared with an exponent of 4.0 for a globular protein, suggesting that NS5FL is less rigid in solution.

Since the crystallographic structure of NS5 (PDB ID: 4V0R) did not contain the residues 1-5 and 896-900 (DENV-3 NS5₆₋₈₉₅) (84), and the construct used in this study (DENV-3 NS5FL) contained flexible-tags at both ends, SAXS data were also collected for DENV-3 NS5₆₋₈₉₅, which was kindly prepared by Dr. Yongqian Zhao (Duke-NUS, Singapore), in order to confirm that the differences seen are due to NS5FL alone. Similar to DENV-3 NS5FL, both the normalized Kratky plot (Fig. 3.2.7A) and Porod-Debye plot (Fig. 3.2.7B) of DENV-3 NS5₆₋₈₉₅ displayed a similar outcome, confirming that the lower rigidity seen in DENV-3 NS5FL is caused by the protein itself, but not caused by the flexible tags and extra residues present in the recombinant DENV-3 NS5FL.

Figure 3.2.7 Assessment of protein flexibility in DENV-3 NS5FL. (A) The normalized Kratky plots for lysozyme (*grey filled circle*), DENV-3 NS5FL (*blue filled circle*) and DENV-3 NS5₆₋₈₉₅ (*orange filled circle*). Lysozyme showed a typical profile of a compact globular protein, while both NS5 proteins showed a profile commonly seen in multi-domain proteins tethered by the linker. (B) The Porod-Debye plots of DENV-3 NS5FL (*blue filled circle*) and DENV-3 NS5₆₋₈₉₅ (*orange filled circle*). The data regions used to compute the Porod exponent were shown as the fitted line for DENV-3 NS5FL (*blue line*) and DENV-3 NS5₆₋₈₉₅ (*orange line*). The reported Porod exponent for both proteins was 3.7, which was smaller from the exponent of 4.0 for a compact globular protein.

To further characterize the potential dynamic behavior in DENV-3 NS5FL, the ensemble optimization method (EOM) was used to analyze the SAXS data of DENV-3 NS5FL. Instead of considering single conformation, EOM analyzes the SAXS data by considering an ensemble of different conformations coexisting in solution. EOM selects the ensemble of conformers that fit to the experimental data and presents the result as a size distribution (R_g and D_{max}) (132,133). Depending on the position of selected R_g distribution with respect to random pool, one can determine whether the

Figure 3.2.8 Ensemble formation of DENV-3 NS5FL. (A) The calculated ensemble scattering profile (red line) fitted well to the experimental scattering profile (black circle), with a discrepancy χ^2 of 0.542. (B) The R_g distribution of selected ensembles from DENV-3 NS5FL exhibited a unimodal distribution, and it was positioned left to the random pool, indicating a single ensemble population of DENV-3 NS5FL with the two domains close to each other. (inset) The arrow indicated the R_g calculated from the crystallographic structure of DENV-3 NS5₆₋₈₉₆ (PDB ID: 4V0R), which has a R_g that is smaller than the R_g from selected ensembles. (C) The representative models for DENV-3 NS5FL from the EOM calculation with the least χ^2 value. The MTase and RdRp were represented as cartoon, and colored in blue and magenta, respectively. The missing residues, which were re-built, were represented as grey spheres.

distance between the domains (inter-domain distance) is close to each other or away from each other. The selected ensemble for DENV-3 NS5FL yielded a discrepancy χ^2 of 0.542 (Fig. 3.2.8A), and the selected representative models were shown in figure 3.2.8C. DENV-3 NS5FL had a unimodal R_g distribution of selected ensembles, with major distribution between 3.2 nm to 3.9 nm (Fig. 3.2.8B) that was less than random pool, showing that DENV-3 NS5FL in solution has a single ensemble population with short inter-domain distance between the two domains of DENV-3 NS5FL. The R_{flex} , which is a quantitative measurement for flexibility, of DENV-3 NS5FL, was 68%, with a randomness threshold of 87% calculated from the random pool, suggesting that the MTase with respect to the RdRp adopted a non-random motion in solution. The quality of the EOM data was confirmed by the R_σ of 0.46, which is expected to be lower than 1.0 when ensemble R_{flex} is less than pool R_{flex} . In comparison, the crystallographic structure of DENV-3 NS5₆₋₈₉₅ (PDB ID: 4V0R) had a smaller R_g

(determined using CRY SOL) being outside the range of selected ensembles. Altogether, the EOM analysis of DENV-3 NS5FL suggests that the movement of the MTase with respect to the RdRp is not random in solution, and with the 10-residue linker tethering two domains, the inter-domain distance between both domains are short, whereby they prefer to stay close to each other as compared to stretch out from each other.

3.2.6 SAXS study of individual domains of DENV-3 NS5FL

SAXS data were collected for DENV-3 MTase₁₋₂₇₂ and RdRp₂₆₃₋₉₀₀, in order to study the contribution of the individual domains to the mobility behavior of DENV-3 NS5FL. The data were collected at protein concentrations of 1.5, 3.3 and 5 mg/ml for the MTase₁₋₂₇₂, and 1.1, 2.1 and 5.56 mg/ml for the RdRp₂₆₃₋₉₀₀. No concentration-dependent effect was identified in the MTase₁₋₂₇₂ sample (Fig. 3.2.9B) and thus, the scattering data with highest concentration (Fig. 3.2.9A), which had less noise, was used for further data analysis. The scattering data of the RdRp₂₆₃₋₉₀₀ were extrapolated to infinite dilution (Fig. 3.2.9D) due to the presence of concentration effect (Fig. 3.2.9E).

Protein sample	MTase ₁₋₂₇₂	RdRp ₂₆₃₋₉₀₀
Concentration range (mg ml ⁻¹)	1.50 – 5.00	1.10 – 5.56
Structural parameters		
$I(0)$ (arbitrary units) [from P(r)]	40.8 ± 0.2	86.8 ± 0.3
R_g (nm) [from P(r)]	2.21 ± 0.01	2.73 ± 0.01
$I(0)$ (arbitrary units) [from Guinier]	40.9 ± 0.3	85.9 ± 0.4
R_g (nm) [from Guinier]	2.20 ± 0.02	2.68 ± 0.02
D_{max} (nm)	7.3 ± 0.7	8.2 ± 0.8
Porod volume estimate (nm ³)	~40.2	~98.5
DAMMIF excluded volume (nm ³)	~54.2	~131
Dry volume (nm ³) [from sequence] ‡	~39	~91
Molecular mass determination		
Calculated monomeric MM (kDa) [from sequence] *	~32.2	~75.3
Molecular mass MM (kDa) [from Porod invariant]	25.1 ± 5	61.6 ± 12
Molecular mass MM (kDa) [from excluded volume]	27 ± 2.7	65 ± 6.5

Software employed

Primary data reduction

BRUKER SAS

Data processing

PRIMUS

* http://web.expasy.org/compute_pi/‡ <http://www.basic.northwestern.edu/biotools/proteincalc.html>**Table 3.2.1** Structural parameters for individual domains of DENV-3 NS5FL

Figure 3.2.9 Primary SAXS data analysis of DENV-3 NS5 domains. (A) SAXS scattering pattern (black circle) of DENV-3 MTase₁₋₂₇₂ at a concentration of 5 mg/ml. The Guinier plot (inset) showed linearity, indicating no protein aggregation. (B) The SAXS scattering patterns of the MTase₁₋₂₇₂ at very low angle overlapped nicely for all three protein concentrations, indicating no inter-particle interaction. (C) $P(r)$ function of DENV-3 MTase₁₋₂₇₂. (D) SAXS scattering pattern (black circle) of DENV-3 RdRp₂₆₃₋₉₀₀ at theoretical infinite dilution. (inset) The Guinier plot showed linearity, indicating no protein aggregation. (E) SAXS scattering patterns of the RdRp₂₆₃₋₉₀₀ at very low angle shifted upward with increasing concentration, indicating the presence of inter-particle interaction. (F) $P(r)$ function of DENV-3 RdRp₂₆₃₋₉₀₀.

The Guinier plots (*inset*) of the MTase₁₋₂₇₂ (Fig. 3.2.9A) and the RdRp₂₆₃₋₉₀₀ (Fig. 3.2.9D) appeared linear, indicating no protein aggregation. The structural parameters for both proteins are shown in table 3.2.1. The R_g determined from Guinier approximation for the MTase₁₋₂₇₂ and the RdRp₂₆₃₋₉₀₀ were 2.2 ± 0.02 nm and 2.68 ± 0.02 nm, respectively. The D_{max} for the MTase₁₋₂₇₂ and the RdRp₂₆₃₋₉₀₀ were 7.3 ± 0.7 nm and 8.2 ± 0.8 nm, respectively.

Comparison of the normalized Kratky plots of DENV-3 MTase₁₋₂₇₂, RdRp₂₆₃₋₉₀₀ and NS5FL (Fig. 3.2.10A) revealed that the bell-curve profile of the MTase₁₋₂₇₂ had a similar shift as NS5FL, showing that the MTase₁₋₂₇₂ is also not rigid in solution. RdRp₂₆₃₋₉₀₀ had a well-defined bell-curve profile as seen in a compact globular protein. In the Porod-Debye plots of all three proteins (Fig. 3.2.10B), RdRp₂₆₃₋₉₀₀ showed a plateau with an exponent of 4.0, indicating that RdRp is a globular compact particle. Similar to the full-length NS5, the MTase₁₋₂₇₂ did not have a plateau (Porod exponent of 3.5), suggesting the changes observed in DENV-3 NS5FL may be mainly contributed by the MTase domain.

Figure 3.2.10 Assessment of protein flexibility in DENV-3 NS5 domains. (A) The normalized Kratky plots for lysozyme (*grey filled circle*), DENV-3 NS5FL (*blue filled circle*), DENV-3 MTase₁₋₂₇₂ (*red filled circle*) and DENV-3 RdRp₂₆₃₋₉₀₀ (*wine filled circle*). Lysozyme and the RdRp₂₆₃₋₉₀₀ showed a typical profile of compact globular proteins, while MTase₁₋₂₇₂ had a profile with a maximum peak shifted close to NS5FL, but a less extended conformation. (B) The Porod-Debye plots of DENV-3 NS5FL (*blue filled circle*), DENV-3 MTase₁₋₂₇₂ (*red filled circle*) and DENV-3 RdRp₂₆₃₋₉₀₀ (*wine filled circle*). The data regions used to compute the Porod exponent were shown as the fitted line for DENV-3 NS5FL (*blue line*), DENV-3 MTase₁₋₂₇₂ (*red line*) and DENV-3 RdRp₂₆₃₋₉₀₀ (*wine line*). The reported Porod exponents for the MTase₁₋₂₇₂ and the RdRp₂₆₃₋₉₀₀ were 3.5 and 4.0, respectively, suggesting that the less rigidity in DENV-3 NS5FL is mainly contributed by the MTase.

3.3 Characterization of NS5FL from DENV-1 to -4 in solution

The multiple sequence alignment of NS5FL showed high protein sequence similarity (75% - 82%) across the four NS5 DENV serotypes (Fig. 3.1.1). In general, high sequence conservation is found in the methyltransferase and RNA-dependent RNA polymerase, despite the fact that some regions within the domain have lower sequence similarity. Previous studies showed that NS5s from DENV-1 to DENV-4 localize at different compartments in the host cells (73,74), where NS5 from DENV-2 and DENV-3 are predominantly localized in the nucleus, and DENV-1 and DENV-4 reside predominantly in the cytoplasm. Up to date, no structural study has been performed for NS5 from all four serotypes. To identify the potential difference in these four proteins, the full-length NS5 proteins from DENV-1 to DENV-4 were studied in solution using SAXS.

3.3.1 Expression and purification of NS5FL from DENV-1 to -4

Induction tests were performed for four NS5FL proteins to ensure the protein production in the *E. coli* cells, and results showed different NS5FL production levels across all four DENV serotypes. Since the protein production of NS5FL from DENV-2 Singapore strain was very low, NS5FL from DENV-2 NGC (New Guinea C) strain was used for the SAXS study. The NS5FL proteins were purified as described in section 2.2.4.1. NS5FL from all four serotypes were eluted from the Ni-NTA column at an imidazole concentration of 50 mM to 150 mM. Impurities were identified by a SDS gel. Afterwards, these samples were concentrated and applied to the SEC column to remove all these impurities and the possible oligomeric forms of NS5FL. Similar to DENV-3 NS5FL, four peaks were identified from the SEC elution profiles of the NS5FL from serotype 1, 2 and 4. NS5FL of all four serotypes were identified at peak III, and the high purity of NS5FL was confirmed by SDS gel (Fig. 3.3.1). The elution fractions corresponding to the top 15% of the protein peaks were pooled and concentrated to three protein concentrations for SAXS experiments.

Figure 3.3.1 The SDS gel showed high purity of NS5FL from DENV-1 to DENV-4.

3.3.2 Primary SAXS data analysis of NS5FL from DENV-1 to -4

Like DENV-3 NS5FL, the SAXS data of NS5FL from DENV-1, DENV-2 and DENV-4 were collected immediately after the last purification step. The buffer scattering (SEC buffer: 50 mM Tris/HCl, pH 7.5, 300 mM NaCl, 5% glycerol, 1 mM DTT) was subtracted from the collected protein data. The molecular masses of all four NS5FL proteins were determined by MALDI-TOF spectrometry, confirming the correct size of all recombinant proteins (Table 3.3.1). The data of three protein concentrations of each protein were recorded, and concentration-dependent effects were identified in all three NS5FL proteins. Therefore, the data of all three NS5FL were extrapolated to infinite dilution (Fig. 3.3.2A) and used for data analysis. The overall structural parameters derived from SAXS experiments are shown in table 3.3.1. The Guinier plots at very low angles showed linearity for all three NS5FL, indicating good data quality for all the samples and no aggregation were seen in all three proteins (*inset* of fig. 3.3.2A). Based on Guinier approximation, the R_g values of NS5FL from DENV-1 to DENV-4 were determined to be 3.55 ± 0.09 nm, 3.56 ± 0.05 nm, 3.37 ± 0.02 nm and 3.27 ± 0.02 nm, respectively. These R_g values were compared with the values derived from the pair-distance distribution ($P(r)$) functions, which were 3.51 ± 0.02 nm, 3.57 ± 0.02 nm, 3.47 ± 0.02 nm and 3.36 ± 0.02 nm, respectively, confirming the R_g values derived from the Guinier approximation.

Both R_g determinations showed that the R_g value of DENV-4 NS5FL was significantly smaller than the NS5 for DENV-1 to -3. The $P(r)$ functions of NS5FL from all four serotypes (Fig. 3.3.2B) were similar, with a profile reflecting slightly elongated conformation. Compared with NS5FL of DENV-1 to DENV-3, the $P(r)$ function of DENV-4 NS5FL shifted slightly inward at the range from 6 nm to 11 nm,

indicating a more compact conformation. The maximum particle dimension D_{max} of all four NS5FLs were determined to be 11 ± 1 nm to 11.5 ± 1 nm.

Figure 3.3.2 Primary SAXS data analysis of DENV-1-4 NS5FL. (A) SAXS scattering patterns of DENV-1 NS5FL (red circle), DENV-2 NS5FL (green circle), DENV-3 NS5FL (blue circle) and DENV-4 NS5FL (magenta circle) at theoretical infinite dilution. The Guinier plots (inset) of all NS5FLs showed linearity, indicating no protein aggregation. (B) $P(r)$ functions of NS5FL from DENV-1 (red line), DENV-2 (green line), DENV-3 (blue line) and DENV-4 (magenta line).

Protein sample	DENV-1 NS5FL	DENV-2 NS5FL	DENV-3 NS5FL	DENV-4 NS5FL
Concentration range (mg ml ⁻¹)	1.17 – 4.56	1.00 – 4.40	1.00 – 4.10	0.45 – 6.86
Structural parameters				
$I(0)$ (arbitrary units) [from $P(r)$]	140.3 ± 0.9	148.9 ± 0.8	137.6 ± 0.7	114.7 ± 0.5
R_g (nm) [from $P(r)$]	3.51 ± 0.02	3.57 ± 0.02	3.47 ± 0.02	3.36 ± 0.02
$I(0)$ (arbitrary units) [from Guinier]	141.9 ± 2.3	149.5 ± 1.5	135.5 ± 1.0	113.0 ± 0.7
R_g (nm) [from Guinier]	3.55 ± 0.1	3.56 ± 0.05	3.37 ± 0.04	3.27 ± 0.03
D_{max} (nm)	11 ± 1	11.5 ± 1	11.5 ± 1	11.5 ± 1
Porod volume estimate (nm ³)	~140	~150	~140	~140
DAMMIF excluded volume (nm ³)	~197	~192	~185	~183
Dry volume (nm ³) [from sequence] ‡	~131	~128	~131	~131
Molecular mass determination				
Calculated monomeric MM (kDa) [from sequence] *	~107.9	~105.3	~107.9	~107.9
Molecular mass MM (kDa)	~108	~105	~108	~108

[from *MALDI*] †

Molecular mass <i>MM</i> (kDa) [from <i>Porod invariant</i>]	87.5 ± 18	93.8 ± 18	87.5 ± 18	87.5 ± 18
Molecular mass <i>MM</i> (kDa) [from <i>excluded volume</i>]	98 ± 10	96 ± 10	93 ± 10	92 ± 10

Software employed

Primary data reduction	BRUKER SAS
Data processing	PRIMUS
<i>Ab initio</i> analysis	DAMMIF & GASBOR
Validation and averaging	DAMAVAR
Computation of model intensities	CRY SOL
Rigid Body Modeling	BUNCH
Flexibility	ScÅtter and EOM 2.0

* http://web.expasy.org/compute_pi/

‡ <http://www.basic.northwestern.edu/biotools/proteincalc.html>

† MALDI-TOF/TOF service provided by Proteomic Core Facility of School of Biological Sciences in Nanyang Technological University

Table 3.3.1 Structural parameters for NS5FL from DENV-1 to -4

3.3.3 Reconstruction of *ab initio* and rigid-body models of NS5FL from DENV-1 to -4

The shapes of NS5FL of DENV-1 to DENV-4 were reconstructed *ab initio* from scattering patterns shown in figure 3.3.2A using program DAMMIF. Twenty independent reconstructed DAMMIF models were computed and the NSD value was calculated between all 20 models using DAMAVAR. The NSD of NS5FL from DENV-1 to DENV-4 were calculated to be 0.756 ± 0.065 , 0.835 ± 0.041 , 0.810 ± 0.053 and 0.777 ± 0.038 , respectively, indicating the reproducible shape reconstructions of all four NS5FL proteins. The averaged models of NS5FL from all four serotypes were shown in figure 3.3.3. All four models have a very similar shape, which is elongated, and a slight bend in the middle of the shape is noted.

The rigid-body models of NS5FL from DENV-1 to -4 were built using BUNCH. Ten independent calculations were performed for each NS5FL, and models with the least χ^2 were superimposed into DAMMIF envelopes as shown in figure 3.3.3. Like for DENV-3 NS5FL, the MTase of the other three serotypes were positioned next to the RdRp in the BUNCH models, and form an elongated conformation.

Figure 3.3.3 *Ab initio* reconstruction and rigid body modeling for NS5FL from DENV-1 to DENV-4. The superimposition between the DAMMIF envelopes (grey sphere) and the BUNCH models (cartoon representation) calculated from the SAXS scattering profiles for NS5FL from all four serotypes. The superimposed structures were shown as front (*top*) and side (*bottom*) views.

3.3.4 Ensemble analysis of NS5FL from DENV-1 to -4

The SAXS curves of NS5FL from all four serotypes were normalized against $I(0)$ and R_g to generate the normalized Kratky plots. Like DENV-3 NS5FL, the normalized Kratky plots of NS5FL from DENV-1, DENV-2 and DENV-4 adopted a right and upward shifted bell-curve profiles compared to the bell-curve profile of globular protein (Fig. 3.3.4A), indicating that NS5FL from all four serotypes were elongated in solution. The bell-curve profile of DENV-4 NS5FL was shifted inward (*red arrow* in Fig. 3.3.4A) when compared with the normalized Kratky plots of the other three serotypes, suggesting that the DENV-4 NS5FL adopted a more compact conformation compared to NS5FL from DENV-1 to -3. The same phenomenon was also seen in the $P(r)$ function (Fig. 3.3.4B). These results suggest that DENV-4 NS5FL adopts a more compact conformation in solution.

Figure 3.3.4 Assessment of protein flexibility in NS5 of DENV-1 to -4. (A) Normalized Kratky plots of NS5FL from DENV-1 (red filled circle), DENV-2 (green filled circle), DENV-3 (blue filled circle) and DENV-4 (magenta filled circle). The lysozyme plot was shown as grey filled circle. The normalized Kratky plot of DENV-4 NS5FL (red arrow) was shifted inward when compared to the NS5FL from other serotypes, suggesting a more compact conformation for the NS5FL in DENV-4. (B) $P(r)$ functions of NS5FL from DENV-1 (red line), DENV-2 (green line), DENV-3 (blue line) and DENV-4 (magenta line). The $P(r)$ function of the DENV-4 NS5FL shifted inward (red arrow) when compared to the NS5FL from the other three serotypes.

The EOM analysis of DENV-3 NS5FL detected a non-random motion between its MTase and RdRp. In order to detect the presence of the same motion in the NS5FL of the other serotypes, the SAXS data of NS5FL from DENV-1, DENV-2 and DENV-4 were subjected to EOM analysis. Similar to DENV-3 NS5FL, NS5FL of serotype 1, 2 and 4 had a unimodal R_g distribution for selected ensembles, as shown in figure 3.3.5. The major R_g distributions of NS5FL covered the range of 3.2-4.0 nm for DENV-1 to DENV-3, and the range of 3.2-3.6 nm for DENV-4, confirming a smaller size of DENV-4 NS5FL compared to the others. The selected R_g distributions of NS5FL from all four serotypes had the maxima less than the random pool, showing a general arrangement of all NS5 proteins with two domains close to each other. The R_{flex} of selected ensembles for NS5FL from DENV-1 to DENV-4 were 81%, 67%, 68% and 74%, respectively, which were smaller than the randomness threshold of 86-88% calculated from the random pool. The quality of the calculations was confirmed with the R_σ of 1.0, 0.43, 0.46 and 0.97, respectively. These EOM data showed that NS5FL from all four serotypes adopt a non-random movement in solution. Altogether, the EOM data analysis of NS5FL suggest that for all four DENV serotypes, with the 10-residue linker tethering two domains, the MTase and the RdRp domains, have short inter-domain distance, whereby they are close to each other, and DENV-4 NS5FL has a more compact domain-domain arrangement compared to others.

Figure 3.3.5 EOM analysis of NS5FL from DENV-1 to -4. The major R_g distributions of NS5FL covered the range of 3.2-4.0 nm for DENV-1 to -3, and the range of 3.2-3.6 nm for DENV-4. The unimodal R_g distributions of NS5FL from all four serotypes had the maxima less than the random pool, showing that NS5FLs have a single ensemble population with the two domains close to each other.

3.4 Mutational studies of the NS5FL linker region of DENV-3 and DENV-4 using SAXS

SAXS studies on all four DENV serotypes reveal a more compact conformation of DENV-4 NS5FL. The MTase and the RdRp of NS5 are connected via a linker, which shows different amino acid composition across the four serotypes (Fig. 3.1.1). To gain more insight into the differences seen in DENV-4 NS5FL and whether they may be caused by the linker, mutational studies were performed using SAXS, by targeting the NS5FL from DENV-3 and DENV-4.

3.4.1 SAXS study of NS5FL linker mutants based on linker substitution

The linker mutational study was performed using NS5FL from DENV-3 and DENV-4. DENV-3 NS5FL was selected as a representative of the NS5FL from DENV-1 to -3. The linker of NS5 from DENV-3 (residues 263 to 272) was swapped to DENV-4 (residues 264 to 273), and *vice versa*, to generate the D3NS5-d4linker and D4NS5-d3linker mutants (Fig. 3.4.1A). Both mutant proteins were produced and purified into high purity samples (Fig. 3.4.1B).

Figure 3.4.1 Protein purification of NS5FL linker mutants. (A) Schematic diagram of NS5FL linker mutants. D3NS5-d4linker (colored as *blue*) and D4NS5-d3NS5 (colored as *magenta*) were generated by swapping the 10-residue linker (SVSTETEKPD) from DENV-4 NS5FL with the DENV-3 NS5FL (HVNAEPETPN), and *vice versa*. (B) The SEC elution

profiles of NS5FL linker mutants. Like wild-type NS5FL, four peaks were identified in the elution profile. The recombinant proteins were identified in the elution fractions corresponding to peak III, and the SDS gels showed high purity of both D3NS5-d4linker (*top inset*) and D4NS5-d3linker (*bottom inset*).

The scattering data of both proteins were collected at three protein concentrations, and the buffer scattering (SEC buffer: 50 mM Tris/HCl, pH 7.5, 300 mM NaCl, 5% glycerol, 1 mM DTT) was subtracted from the collected protein data. Concentration-dependent effects were observed. Thus, the scattering data with extrapolation to infinite dilution were used for further data analysis. No protein aggregation was seen, as shown in the linearity of Guinier plots (*inset* of Fig. 3.4.2). The derived R_g and D_{max} for the D3NS5-d4linker and D4NS5-d3linker were 3.58 ± 0.02 nm and 11.0 ± 1 nm, and 3.26 ± 0.01 nm and 10.0 ± 1 nm, respectively. The overall structural parameters for both mutants are shown in table 3.4.1.

Protein sample	D3NS5-d4linker	D4NS5-d3linker
Concentration range (mg ml ⁻¹)	1.00 – 3.10	1.90 – 6.30
Structural parameters		
<i>I</i> (0) (arbitrary units) [from P(r)]	143.4 ± 0.8	104.7 ± 0.4
<i>R_g</i> (nm) [from P(r)]	3.58 ± 0.02	3.26 ± 0.01
<i>I</i> (0) (arbitrary units) [from Guinier]	144.2 ± 1.2	102.8 ± 0.3
<i>R_g</i> (nm) [from Guinier]	3.57 ± 0.05	3.17 ± 0.03
<i>D_{max}</i> (nm)	11.0 ± 1	10.0 ± 1
Porod volume estimate (nm ³)	~147	~130
DAMMIF excluded volume (nm ³)	~197	~179
Dry volume (nm ³) [from sequence] ‡	~131	~131
Molecular mass determination		
Calculated monomeric <i>MM</i> (kDa) [from sequence] *	~107.9	~107.9
Molecular mass <i>MM</i> (kDa) [from <i>Porod invariant</i>]	91.1 ± 18	81.3 ± 16
Molecular mass <i>MM</i> (kDa) [from <i>excluded volume</i>]	99 ± 10	90 ± 9
Software employed		
Primary data reduction	BRUKER SAS	
Data processing	PRIMUS	
* http://web.expasy.org/compute_pi/		
‡ http://www.basic.northwestern.edu/biotools/proteincalc.html		

Table 3.4.1 Structural parameters for NS5FL mutants D3NS5-d4linker and D4NS5-d3linker

Figure 3.4.2 SAXS scattering patterns of NS5FL linker mutants: D3NS5-d4linker (blue circle) and D4NS5-d3linker (magenta circle) at theoretical infinite dilution. The Guinier plots (inset) of both proteins showed linearity, indicating no protein aggregation.

Figure 3.4.3 SAXS analysis of NS5FL linker mutants. (A) $P(r)$ functions of wild-type DENV-3 NS5FL (blue solid line) and DENV-4 NS5FL (magenta solid line), and mutant D3NS5-d4linker (blue dashed line) and D4NS5-d3NS5 (magenta dashed line). (B) Normalized Kratky plots of lysozyme (grey filled circle), wild-type DENV-3 NS5FL (blue filled circle) and DENV-4 NS5FL (magenta filled circle), and linker mutant D3NS5-d4linker (blue circle) and D4NS5-d3linker (magenta circle). The plot of wild-type DENV-3 NS5FL overlaid nicely with the plot of the mutant D3NS5-d4linker. The normalized Kratky plot of D4NS5-d3linker shifted inward significantly compared to wild-type DENV-4 NS5FL, indicating a change in compactness after introducing the linker residues from DENV-3 NS5FL into DENV-4 NS5FL.

Comparing D3NS5-d4linker with DENV-3 NS5FL, the R_g increased slightly and the $P(r)$ function (Fig. 3.4.3A) shifted slightly outward, but no significant difference was observed in the normalized Kratky plots (Fig. 3.4.3B), suggesting that the introduction of DENV-4 NS5 linker into DENV-3 NS5 has slight or no effect to the flexibility of DENV-3 NS5. While in the case of the D4NS5-d3linker, swapping the DENV-3 NS5 linker into DENV-4 NS5 caused a more compact conformation as seen from the inward shift in both $P(r)$ functions (Fig. 3.4.3A) and normalized Kratky plots (Fig. 3.4.3B). These results suggest that the change of linker region from DENV-3 to

DENV-4 NS5 affects its flexibility, and the linker of DENV-4 NS5FL is essential for maintaining proper compactness of DENV-4 NS5.

3.4.2 NS5 linker mutant design using amino acid substitution

To further identify the residue(s) responsible for the difference seen in the D3NS5-d4linker and D4NS5-d3linker mutants, seven DENV-4 NS5 mutants were created, and the mutated amino acid sequences at the linker region are shown in table 3.4.2.

Protein sample	Linker sequences
DENV-4 NS5FL	SVSTETEKPD
DENV-4 NS5 S264H	HVSTETEKPD
DENV-4 NS5 K271T	SVSTETETPD
DENV-4 NS5 K271T/D273N	SVSTETETPN
DENV-4 NS5 T269N/K271T/D273N	SVSTENETPN
DENV-4 NS5 S264H/T269N/K271T/D273N	HVSTENETPN
DENV-4 NS5 S266N/T267A	SVNAETEKPD
DENV-4 NS5 Δ S ₂₆₆ T ₂₆₇	SVSTETEKPD

Table 3.4.2 Amino acid sequences of DENV-4 NS5FL linker mutants

The comparison of linker amino acid sequences between DENV-4 NS5 and DENV-1-3 NS5 is shown in figure 3.4.4A. S264, S266, T267, T269, K271 and D273 from DENV-4 NS5 were selected and replaced with their corresponding linker residues found in DENV-3 NS5. The mutated residues were selected based on the non-conserved residues across four serotypes, and their interacting partner between linker residues and domains' residues (Fig. 3.4.4B). S264H mutant was created because from the DENV-3 NS5₆₋₈₉₅ crystallographic structure, H263, which corresponds to S264 in DENV-4 NS5, interacts with T261 from the MTase domain, and it is not highly conserved among four serotypes. K271T mutant was designed because of the charge and size properties of lysine and its corresponding residue in other three serotypes is threonine/valine. Based on these two mutants, three more mutants with more than one mutation sites had been created, and all the mutated residues are not conserved, and they were K271T/D273N, T269P/K271T/D273N and S264H/T269P/K271T/D273N. Besides these five mutants, two residues from DENV-4 NS5, S266 and T267, are not conserved among the four serotypes. Comparing the amino acid properties of these residues, only DENV-4 NS5 has two polar residues in this region, while in other three serotypes, at least one of the residues is found to be non-polar. Therefore, two mutants,

S266N/T267A and Δ S₂₆₆T₂₆₇ (deletion of S266 and T267) were designed, to understand the role of these two residues.

Figure 3.4.4 Sequence analysis of NS5FL linker mutants. (A) The comparison of NS5 linker residues from DENV-1 to DENV-4. The conserved residues are highlighted as *purple*. The linker residues interacting with the NS5 domains are identified from the DENV-3 NS5₆₋₈₉₅ crystallographic structure (PDB ID: 4V0R) (84) and colored as *blue*. (B) Interacting residues between linker region (*green cartoon representation*) and the MTase (*blue cartoon representation*). Two linker residues (*green stick representation*), H263 and E267 interact with the MTase residues (*blue stick representation*), T261 and R262, respectively. (C) Interacting residues between linker region (*green cartoon representation*) and RdRp (*magenta cartoon representation*). Three linker residues (*green stick representation*), E269, T270 and N272 interact with the RdRp residues (*magenta stick representation*), K595 and N574.

3.4.3 SAXS study of NS5FL linker mutants based on amino acid substitution

All the NS5FL linker mutant constructs were cloned by Dr. Ardina Grüber and Ms. Ankita Pan from our laboratory. These mutant proteins were purified into high purity samples. For each mutant protein, the SAXS scattering profiles were collected at three protein concentrations, and concentration dependency was analyzed at the low angle region of the scattering curves. Similar to wild-type DENV-4 NS5, concentration effects were observed in all mutant proteins. Therefore, the collected data were extrapolated to theoretical infinite dilution for data analysis. Observing the Guinier region of all three protein concentrations from each mutant protein, linearity of the Guinier plots confirmed good data quality and no protein aggregation. Further data analyses for all seven mutants were done by Dr. Malathy Sony Subramanian

Manimekalai from our laboratory. In short, the derived R_g for all DENV-4 NS5 mutant proteins based on the Guinier approximation and the $P(r)$ function, and maximum particle dimension (D_{max}) of the protein, were shown in table 3.4.3, together with the values from wild-type DENV-3 NS5FL and DENV-4 NS5FL.

Protein sample	R_g (nm) (from Guinier)	R_g (nm) (from $P(r)$)	D_{max} (nm) (from $P(r)$)
DENV-3 NS5FL	3.37 ± 0.04	3.47 ± 0.02	11.5 ± 1
DENV-4 NS5FL	3.27 ± 0.03	3.36 ± 0.02	11.5 ± 1
DENV-4 NS5 S264H	3.28 ± 0.04	3.39 ± 0.02	10.7 ± 1
DENV-4 NS5 K271T	3.43 ± 0.03	3.45 ± 0.02	11.1 ± 1
DENV-4 NS5 K271T/D273N	3.34 ± 0.03	3.39 ± 0.02	10.9 ± 1
DENV-4 NS5 T269N/K271T/D273N	3.32 ± 0.04	3.38 ± 0.02	11.0 ± 1
DENV-4 NS5 S264H/T269N/K271T/D273N	3.26 ± 0.03	3.37 ± 0.03	10.9 ± 1
DENV-4 NS5 S266N/T267A	3.42 ± 0.04	3.49 ± 0.02	11.1 ± 1
DENV-4 NS5 ΔS₂₆₆T₂₆₇	3.44 ± 0.05	3.48 ± 0.02	11.1 ± 1

Table 3.4.3 Scattering-derived parameters from Guinier region and $P(r)$ function for wild-type NS5 and DENV-4 NS5FL linker mutants

Comparing the dimensions among wild-type DENV-4 NS5 and its mutants, the DENV-4 NS5 K271T, S266N/T267A and Δ S₂₆₆T₂₆₇ mutants reveal a significant increase in their R_g . To further understand the difference between these three mutants and other mutants, pair-distance distribution ($P(r)$) function and normalized Kratky plots for all mutants were created and shown in figure 3.4.5. The NS5 mutant S264H, K271T/D273N, T269N/K271T/D273N and S264H/T269N/K271T/D273N had similar $P(r)$ functions (Fig. 3.4.5A) and normalized Kratky plots (Fig. 3.4.5C) as compared to wild-type DENV-4 NS5FL. Compared to the NS5 mutants which showed larger R_g values (K271T, S266N/T267A and Δ S₂₆₆T₂₆₇), more long distance pairs were found in these three mutants, as indicated from the outward shift of their $P(r)$ functions at the tail region (*red arrow* of Fig. 3.4.5B). Besides this, their normalized Kratky plots also shifted outward when compared with the DENV-4 NS5FL (Fig. 3.4.5D), indicating an elongation and increase of the flexibility in these three proteins compared to wild-type DENV-4 NS5 and the other four mutants. For comparison, the $P(r)$ function and the normalized Kratky plot from DENV-3 NS5FL were also included in all figures.

Figure 3.4.5 SAXS analysis of DENV-4 NS5FL linker mutants. (A, B) $P(r)$ functions (solid line) and (C, D) normalized Kratky plots (filled circle) of wild-type DENV-3 NS5FL (blue) and DENV-4 NS5FL (magenta), and DENV-4 NS5FL mutants. (A) The $P(r)$ functions and (C) normalized Kratky plots of mutants S264H (cyan), K271T/D273N (green), T269N/K271T/D273N (purple) and S264H/T269N/K271T/D273N (wine) overlaid nicely (red arrow) to the wild-type DENV-4 NS5FL, while (B) the $P(r)$ functions and (D) normalized Kratky plots of mutants K271T (violet), S266N/T267A (dark yellow) and $\Delta S_{266}T_{267}$ (orange) shifted away (red arrow) from DENV-4 NS5FL and toward DENV-3 NS5FL. These figures are adapted from Subramanian Manimekalai *et al.*, 2016 (176).

3.4.4 Ensemble analysis of NS5FL mutants based on amino acid substitution

From the primary SAXS data analysis of these linker mutants, differences in flexibility were identified. To further analyze the flexibility among these seven mutants, EOM was performed using the same strategy for wild-type NS5. This work was done by Dr. Malathy Sony Subramanian Manimekalai. In brief, from the EOM analysis, the comparison of the R_g distributions of all seven mutants with the wild-type DENV-4 NS5FL revealed that the mutants K271T, S266N/T267A and $\Delta S_{266}T_{267}$ had a slight change in their R_g distributions. The R_g distributions of mutants S264H, K271T/D273N, T269N/K271T/D273N and S264H/T269N/ K271T/D273N overlapped nicely with the wild-type NS5, as shown in figure 3.4.6A. In contrast, in NS5 mutants K271T, S266N/T267A and $\Delta S_{266}T_{267}$ the distribution at lower R_g (~ 3.3 nm; blue arrow in Fig. 3.4.6B) decreased, while the distribution increased at higher R_g values (~ 3.7 nm; red arrow in Fig. 3.4.6B) has increased. The selected conformations by EOM

from wild-type NS5 and all NS5 mutants showed one major conformation, which had an R_g of 3.4 nm and D_{max} of 12 nm (red boxes in Fig. 3.4.6C and D), and a few minor conformations. It is clearly shown that the major conformation of NS5 mutants S264H (~80%), K271T/D273N (~88%), T269N/K271T/D273N (~78%) and S264H/T269N/K271T/D273N (~90%) have a similar population as DENV-4 NS5FL (~73%), but the mutants K271T (~50%), S266N/T267A (~50%) and $\Delta S_{266}T_{267}$ (~50%) have a similar population as DENV-3 NS5FL (~50%). Altogether, these results suggest that the K271, S266N and T267 are critical residues of the linker region being responsible for the compactness observed in DENV-4 NS5FL.

Figure 3.4.6 Ensemble analysis of DENV-4 NS5FL linker mutants. (A, B) The comparison of R_g distributions of selected ensemble from wild-type DENV-4 NS5FL (magenta) and the mutants, with the random pool (grey shaded area). The R_g distribution from (A) S264H (cyan), K271T/D273N (green), T269N/K271T/D273N (purple) and S264H/T269N/K271T/D273N (wine) did not show significant difference compared to wild-type NS5FL, while (B) the R_g distribution at ~3.3 nm (blue arrow) of K271T (violet), S266N/T267A (dark yellow) and $\Delta S_{266}T_{267}$ (orange) decreased, and there was another R_g distribution at ~3.7 nm (red arrow) in these three mutants. These figures are adapted from Subramanian Manimekalai *et al.*, 2016 (176). (C, D) The populations of the selected conformers from EOM for wild-type (filled circle) DENV-3 NS5FL (blue) and DENV-4 NS5FL (magenta), and DENV-4 NS5FL mutants (filled square). The mutants (C) S264H (cyan), K271T/D273N (green), T269N/K271T/D273N (purple) and S264H/T269N/K271T/D273N (wine) had the similar populations of major conformations compared to DENV-4 NS5FL, while (D) K271T (violet), S266N/T267A (dark yellow) and $\Delta S_{266}T_{267}$ (orange) had similar populations of major conformations as DENV-3 NS5FL.

3.5 Characterization of full-length NS5 (NS5FL) with substrates using SAXS

NS5 is composed of two active enzymes which catalyze important enzymatic steps in viral replication. Its MTase domain is known to catalyze the genomic RNA capping, and its RdRp domain is responsible for the synthesis of new genomic viral RNA. In order to carry out their enzymatic activities, both enzymes require different substrates and cofactors. Conformational changes may arise during the binding, catalytic and releasing events of these substrates and cofactors. Therefore, in order to observe any change that may occur in NS5 upon substrate binding or releasing, SAXS studies for the NS5 protein in the absence and presence of substrates were carried out using the synchrotron light source.

3.5.1 Static SAXS studies of DENV-2 NS5FL with substrates

GTP (Guanosine triphosphate) is required by the MTase domain of NS5 to generate the 5' cap, and SAM (*S*-adenosyl-methionine) provides the methyl group to the Gppp-cap and release SAH (*S*-adenosyl-homocysteine) as the by-product. Therefore, four substrates: GTP (Guanosine triphosphate), GDP (Guanosine diphosphate), SAM (*S*-adenosyl-methionine) and SAH (*S*-adenosyl-homocysteine) were selected to study the protein-substrate interaction event. The first two nucleotides from viral RNA, A₁ and G₂, are conserved among flaviviruses and contributed significantly to the RNA-binding with MTase (83). Thus, two short RNAs: Gppp-RNA (G₀pppA₁G₂U₃U₄G₅U₆U₇) and cap-0 RNA (m⁷G₀pppA₁G₂U₃U₄G₅U₆U₇), which contain both important nucleotides, were selected in this study. These two RNAs were kindly provided by Novartis (Singapore). Since the interacting event between protein and substrate may only trigger a small change, to capture this small difference, high-resolution SAXS data were collected using a high brilliant synchrotron light source. SAXS data of DENV-2 NS5FL with these substrates were collected. Each sample was recorded at two protein concentrations, 1.25 mg/ml and 2.5 mg/ml. For small molecules (GTP, GDP, SAM, SAH), the data were collected in a protein to substrate molar ratio of 1:3, while for RNA (Gppp-RNA and cap0-RNA), the data were collected in equal molar ratio for protein to RNAs. The scattering of buffer 50 mM Tris/HCl, pH 7.5, 300 mM NaCl, 5% glycerol and 1 mM DTT, with equal amount of substrate that was added to the protein sample, was subtracted from each respective protein-substrate data. The scattering curves from two concentrations were merged together (Fig. 3.5.1A and 3.5.1C) to further improve the quality of scattering data for

data analysis. The summary of scattering-derived parameters for all nine samples is shown in table 3.5.1.

Figure 3.5.1 Primary SAXS analysis of DENV-2 NS5FL in the absence or presence of substrates. (A, C) Scattering curves (*scatter*) and (B, D) the Guinier plots of DENV-2 NS5FL in the absence (apo; *green*) or presence of substrates, including (A, B) guanosine triphosphate (GTP; *red*), guanosine diphosphate (GDP; *blue*), *S*-adenosyl-methionine (SAM; *magenta*), *S*-adenosyl homocysteine (SAH; *dark yellow*), GDP and SAM (*dark cyan*), and GDP and SAH (*orange*), (C, D) Gppp-RNA (*purple*) and cap-0 RNA (*violet*). The Guinier plots derived from all the data showed linearity, indicating no protein aggregation.

DENV-2 NS5FL	R_g (nm) (from Guinier)	R_g (nm) (from $P(r)$)	D_{max} (nm) (from $P(r)$)
apo	3.62 ± 0.01	3.69 ± 0.01	12.3
+ GTP	3.64 ± 0.01	3.68 ± 0.01	12.2
+ GDP	3.65 ± 0.01	3.68 ± 0.01	12.3
+ Gppp-RNA	3.71 ± 0.01	3.78 ± 0.01	13.0
+ cap-0 RNA	3.73 ± 0.01	3.79 ± 0.01	13.0
+ SAM	3.63 ± 0.01	3.66 ± 0.01	12.1
+ SAH	3.57 ± 0.01	3.62 ± 0.01	11.5
+ GDP & SAM	3.60 ± 0.01	3.64 ± 0.01	12.0
+ GDP & SAH	3.61 ± 0.01	3.66 ± 0.01	12.1

Table 3.5.1 Scattering-derived parameters from Guinier region and $P(r)$ functions for DENV-2 NS5FL in the absence or presence of substrates

Figure 3.5.2 SAXS analysis of DENV-2 NS5FL in the absence or presence of substrates. (A, B) $P(r)$ functions (solid line) and (C, D) normalized Kratky plots (filled circle) of DENV-2 NS5FL in the absence (apo; green) or presence of substrates. (A, C) No significant difference (red arrow) was seen in both plots between apo form and in the presence of guanosine triphosphate (GTP; red), guanosine diphosphate (GDP; blue), *S*-adenosyl-methionine (SAM; magenta), *S*-adenosyl homocysteine (SAH; dark yellow), GDP and SAM (dark cyan), and GDP and SAH (orange). (B, D) In the presence of Gppp-RNA (purple) and cap-0 RNA (violet), their $P(r)$ functions (red arrow) and normalized Kratky plots (blue arrow) shifted slightly outward.

Without the substrates (apo form), DENV-2 NS5FL has the R_g value of 3.62 ± 0.01 nm derived from the Guinier approximation, and 3.69 ± 0.01 nm derived from the $P(r)$ function. In the presence of substrates GTP, GDP, SAM and SAH, as well as adding two substrates at once (GDP and SAM or GDP and SAH), the R_g values derived from Guinier approximation and the $P(r)$ function were ranged from 3.57 to 3.65 nm, and 3.62 to 3.68 nm, respectively. As compared to the apo form, no significant R_g difference was seen in these six experiments (Table 3.5.1). The D_{max} values derived from the $P(r)$ functions for these six experiments also showed no significant difference compared to the apo form. However, in the presence of RNA, the R_g and D_{max} values showed significant increase as compared with the apo form. The R_g values derived from the Guinier approximation increased after mixing NS5FL with Gppp-RNA or cap-0 RNA, resulting in the R_g values of 3.71 ± 0.01 nm and 3.73 ± 0.01 nm, respectively. A similar increase was also observed for the R_g values derived from the $P(r)$ function, indicating that the interaction of RNA and the NS5FL protein triggers an increase in the NS5 particle size.

To further assess this difference, the $P(r)$ functions and the normalized Kratky plots for all samples were generated and compared, as shown in figure 3.5.2. The outward shift of the $P(r)$ functions of DENV-2 NS5FL in the presence of Gppp-RNA and cap-0 RNA at longer distance pairs (*red arrow* in Fig. 3.5.2B), and also in normalized Kratky plots (*blue arrow* in Fig. 3.5.2D) indicated that the particles became slightly elongated and more flexible when the RNAs were added into protein solution.

3.5.2 Time-dependent SAXS studies of DENV-4 NS5FL with RNA

From the static SAXS experiment, the particle dimensional parameters of DENV-2 NS5FL were changed in the presence of short capped-RNA. In order to get further insight into this phenomenon, time-dependent SAXS was performed, to identify the time point where the changes occur in the NS5FL protein in the presence of RNA. Even though DENV-2 NS5FL can be purified to a reasonable quantity for the time-dependent SAXS study, it was found out that during the SAXS sample preparation at the synchrotron outstation, due to the freezing and thawing processes, some amounts of the protein were lost because of protein precipitation. Compared to DENV-2 NS5FL, DENV-4 NS5FL can be purified to a reasonable amount, and preserves its stability after freezing and thawing. Therefore, the time-dependent SAXS was performed using DENV-4 NS5FL. Since large amounts of samples are required to perform time-dependent SAXS, a different RNA molecule was used in the study because of the limitation in synthesizing large quantities of capped-RNA. The RNA molecules with the same sequences as the RNA used in static SAXS were purchased from IDT (Integrated DNA Technologies, USA). It was previously shown that RNA with one phosphate group binds to the MTase (177). To increase the binding affinity, a phosphate group was attached to the 5' end of RNA. To allow the saturation of RNA-binding to the protein, a protein to RNA molar ratio of 1:5 was used in time-dependent SAXS. Since different protein and RNA molecule were used compared to previous static SAXS experiment, the static SAXS experiment was repeated with the new samples, prior to the time-dependent SAXS data collection. From the previous SAXS experiment, as concentration dependency was observed in the NS5FL from all serotypes, the data were collected at two protein concentrations, 2.3 mg/ml and 4.5 mg/ml, and in the absence (apo) or presence (+ phosRNA) of RNA. The scattering profiles for all four samples are shown in figure 3.5.3A and their Guinier plots (*inset* of Fig. 3.5.3A) showed linearity, indicating no protein aggregation. The R_g values for

NS5FL in the absence or presence of RNA are 3.83 ± 0.01 nm and 3.94 ± 0.01 nm, respectively at protein concentration of 2.3 mg/ml, and 3.99 ± 0.01 nm and 4.27 ± 0.01 nm, respectively at protein concentration of 4.5 mg/ml. The R_g comparison showed a significant enlargement of NS5FL after mixing with RNA, irrespective of the protein concentration. The same phenomenon was seen in their $P(r)$ functions (Fig. 3.5.3B) and normalized Kratky plots (Fig. 3.5.3C), where the tail of the +phosRNA curve shifted outward compared to the one of apo-NS5FL. These results indicate that RNA induces a significant change in NS5FL.

Figure 3.5.3 Static SAXS analysis of DENV-4 NS5FL in the absence or presence of RNA in two different protein concentrations. (A) SAXS scattering patterns of DENV-4 NS5FL at 2.3 mg/ml (filled circle) and 4.5 mg/ml (empty circle) in the absence (apo) or presence of phosphorylated RNA (+phosRNA), at molar ratio of 1:5 (protein:phosRNA). (top inset) The Guinier plots of all data showed linearity, indicating no protein aggregation. (bottom inset) The scattering intensity (filled circle) of complex (+phosRNA, blue) is significantly higher than of the substrate-free protein (green) due to high electron density of RNA. The sum (grey) of scattering intensities of the protein only and phosRNA only (light magenta) is significantly different from the complex's scattering intensity, indicating the complex formation. (B) At protein concentration of 2.3 mg/ml, the $P(r)$ functions (line) of apo (orange) and +phosRNA (magenta) did not show a significant difference, while at protein concentration of 4.5 mg/ml, the $P(r)$ function shifted outward in the presence of phosRNA (blue). (C) In the normalized Kratky plots (filled circle) of NS5FL, at protein concentration of 2.3 mg/ml, the plot shifted outward in the presence of phosRNA (magenta), as compared to the apo form (orange). Similar trend was observed at protein concentration of 4.5 mg/ml, where in the presence of phosRNA (blue), the plot shifted outward as compared to apo form (green).

To ensure the observed change is not due to the scattering density difference of protein and RNA, the scattering intensities from NS5FL only and phosRNA only were summed together, and compared with the scattering intensity of NS5FL in the presence of RNA (bottom inset of fig. 3.5.3A). Both scattering profiles appeared to be different, supporting that the scattering intensity change is due to the complex formation between NS5FL and RNA. Since a significant change was observed after adding RNA into the NS5FL in the static SAXS experiment, the time-dependent SAXS experiment was performed. The data were collected at two protein concentrations, 4.5 mg/ml and 2.3 mg/ml, and two protein-to-substrate ratios of 1:5 and 1:2. The first data was collected at 1 min, and later on in 5 min interval. The scattering curves for each time point are shown in figure 3.5.4A and 3.5.5A. At protein concentration of 4.5 mg/ml, the R_g values at different time point showed a slight increasing tendency (Fig. 3.5.5B), regardless of the ratios of protein to RNA (1:5 or 1:2). With the increasing amount of RNA (1:5), the size of NS5FL had a significant increase as compared to the ratio of 1:2. With the protein concentration of 2.3 mg/ml, adding RNA triggered a similar trend of increase in the R_g value, as shown in figure 3.5.4B.

Figure 3.5.4 Time-dependent SAXS analysis of DENV-4 NS5FL in the absence or presence of RNA at protein concentration of 2.3 mg/ml. (A) Scattering patterns of DENV-4 NS5FL (line) collected in the presence of phosphorylated RNA, at a protein concentration of 2.3 mg/ml using a protein-to-substrate molar ratio of 1:5. (B) The R_g comparison of the scattering data collected in the presence of phosRNA at different time point, for a protein concentration of 2.3 mg/ml and a protein-to-substrate molar ratio of 1:5 (magenta filled circle). The R_g value from static SAXS experiments in the absence of phosRNA was shown as the grey dashed line.

Figure 3.5.5 Time-dependent SAXS analysis of DENV-4 NS5FL in the absence or presence of RNA at protein concentration of 4.5 mg/ml. (A) Scattering patterns of DENV-4 NS5FL (filled circle) collected at a 5 min intervals for 30 min, in the presence of phosphorylated RNA, at protein concentration of 4.5 mg/ml using protein-to-substrate molar ratio of (left) 1:5 and (right) 1:2. (B) The R_g comparison of the scattering data collected in the presence of phosRNA at different time points, for a protein concentration of 4.5 mg/ml and a protein-to-substrate molar ratio of 1:5 (blue filled circle) and 1:2 (wine filled circle). The R_g value from static SAXS experiments in the absence of phosRNA was shown as the grey dashed line.

Similar to the static SAXS experiment, the time-dependent SAXS experiment (5 min intervals) showed a significant change after adding RNA into NS5FL, especially at protein concentration of 4.5 mg/ml. However, the R_g change was already observed after one minute. Therefore, it was hypothesized that the interacting event between NS5FL and the RNA was a fast event. In order to analyze this fast event, TR-SAXS was performed using a stopped-flow device. The data were collected for each sample in the duration of 50 x 20 ms (In a total of 1 min). The scattering curves from each time point are shown in figure 3.5.6A and 3.5.6B. From each curve, the R_g was calculated and plotted. As shown in figure 3.5.6C, after adding RNA, a slight increase in the R_g value was observed, and the R_g values of averaged data of apo and

+phosRNA were 3.79 ± 0.01 nm and 3.84 ± 0.01 nm, respectively. However, no significant change of R_g value was identified at each time point after adding phosRNA. Taking these results together, the interaction of NS5FL and phosRNA induces an increase of the NS5 particle size, but more experiments are needed to pinpoint the exact time point where the interaction occurs.

Figure 3.5.6 TR-SAXS analysis of DENV-4 NS5FL in the absence or presence of RNA. (A, B) Scattering patterns of DENV-4 NS5FL (filled circle) collected for duration of 50 x 20 ms in (A) the absence (apo; green) or (B) presence (+phosRNA; orange) of phosphorylated RNA. (C) The R_g comparison of the scattering data collected in absence or presence of phosRNA at different time points, for duration of 50 x 20 ms.

3.6 Characterization of DENV-4 helicase in solution using SAXS

Dengue viruses replicate their viral genome in the replication complexes (RCs) comprised of viral RNA, replication proteins and host proteins, located in an invaginated double-membrane vesicle on the endoplasmic reticulum. The viral proteins NS3 (residues 1 – 618) and NS5 form the catalytic core in the RC, and both of them are multifunctional proteins with two domains. NS3 and NS5 were shown to interact with each other (69) through the interaction between NS3 helicase domain (residues 179-618) and NS5 RdRp domain (residues 272-900), where the interacting site was mapped to the C-terminus of the helicase (residues 303-618) and bNLS region of the RdRp (residues 320-368) (71,90). Based on the data from surface plasma resonance (SPR) and NS5 peptide-phage ELISA, the interacting residues from NS5 were further mapped to residues 320-341 (91). To gain some structural insights into NS3-NS5 interaction, SAXS was performed on the DENV helicase (NS3₁₇₂₋₆₁₈) and a fused protein (NS3₁₇₂₋₆₁₈-NS5₃₂₀₋₃₄₁), which contained the NS3 helicase domain and an NS5 peptide spanning residues 320 to 341 (Fig. 3.6.1).

Figure 3.6.1 Schematic diagram of DENV-4 NS3 helicase constructs. The helicase has the NS3 amino acids from 172 to 618. The fused protein, hel-NS5₃₂₀₋₃₄₁ has the NS3 amino acids from 172 to 618, and after residue 618, it is linked to a peptide that contains NS5 amino acids from 320-341, via a G₄SG₄ linker.

3.6.1 Expression and purification of DENV-4 helicase and hel-NS5₃₂₀₋₃₄₁

The DENV-4 helicase (NS3₁₇₂₋₆₁₈) and hel-NS5₃₂₀₋₃₄₁ were purified as described in section 2.2.4.2 (Materials and Methods). Both helicase and hel-NS5₃₂₀₋₃₄₁ were eluted from the Ni-NTA column at the imidazole concentration of 75 mM to 150 mM, as shown in the SDS gel of figure 3.6.2A. To remove the thioredoxin-His₆ tag from the helicase and His₆ tag from hel-NS5₃₂₀₋₃₄₁, these fractions were pooled together and dialyzed overnight with thrombin and TEV protease, respectively. The proteolytically cleaved samples were then subjected to Ni-NTA column again and the cleaved products, namely the helicase and the hel-NS5₃₂₀₋₃₄₁ without any tags, appeared in the flow through (Fig. 3.6.2B). These samples were concentrated and applied to SEC column, and three peaks were identified from the SEC elution profile (Fig. 3.6.2C). Both proteins were identified in the elution fractions corresponding to peak III, where

the helicase was eluted at 16 ml, while the hel-NS5₃₂₀₋₃₄₁ was eluted slightly earlier, which was at 15.7 ml, due to the presence of the small NS5 peptide that fused to the helicase. The elution fractions corresponding to the top 15% of the protein peak (Fig. 3.6.2C) were pooled together and concentrated for SAXS experiment. As shown in the SDS gel (*inset* of fig. 3.6.2C), both proteins were highly pure.

Figure 3.6.2 Purification of DENV-4 helicase and hel-NS5₃₂₀₋₃₄₁ proteins. (A) SDS gel after Ni-NTA affinity chromatography for (*top*) helicase and (*bottom*) hel-NS5₃₂₀₋₃₄₁. The samples from 75-150 mM imidazole fractions were pooled and subjected to cleavage overnight with thrombin or TEV protease to remove the affinity tags. (B) SDS gel after overnight cleavage. The cleaved helicase and hel-NS5₃₂₀₋₃₄₁ were separated nicely from the un-cleaved proteins after a second Ni-NTA affinity column. (C) The SEC elution profile of DENV-4 helicase (*black solid line*) and hel-NS5₃₂₀₋₃₄₁ (*black dashed line*). Three peaks were identified in the elution profile, and both proteins were identified in the elution fractions corresponding to peak III. (*left inset*) SDS gel loaded with the fractions from peak III, and this peak contained high purity of helicase and hel-NS5₃₂₀₋₃₄₁.

3.6.2 Primary SAXS data analysis of DENV-4 helicase and hel-NS5₃₂₀₋₃₄₁

After the last purification step, the DENV-4 helicase and hel-NS5₃₂₀₋₃₄₁ were concentrated to protein concentrations of 1.2, 2.8 and 3.7 mg/ml for the helicase, and 1.2, 3.0 and 4.0 mg/ml for the hel-NS5₃₂₀₋₃₄₁, and used for SAXS data collection. The buffer scattering (buffer C: 20 mM Tris/HCl, pH 7.5, 200 mM NaCl, 1 mM DTT) was subtracted from all collected protein data. The low angle scattering data of both proteins overlapped nicely for all three concentrations, indicating no inter-particle interaction. The Guinier plots of both proteins showed linearity at all three concentrations, showing no protein aggregation. Therefore, the scattering data at concentration of 2.8 mg/ml for the helicase and 3.0 mg/ml for the hel-NS5₃₂₀₋₃₄₁ were used for further data analysis (Fig. 3.6.3A). The overall SAXS structural parameters are shown in table 3.6.1.

Protein sample	Helicase	Hel-NS5 ₃₂₀₋₃₄₁
Concentration range (mg ml ⁻¹)	1.20 – 3.80	1.20 – 4.00
Structural parameters		
$I(0)$ (arbitrary units) [from P(r)]	62.1 ± 0.3	66.3 ± 0.3
R_g (nm) [from P(r)]	2.41 ± 0.01	2.47 ± 0.01
$I(0)$ (arbitrary units) [from Guinier]	63.3 ± 0.4	67.5 ± 0.4
R_g (nm) [from Guinier]	2.50 ± 0.02	2.54 ± 0.03
D_{max} (nm)	7.3 ± 0.7	7.6 ± 0.8
Porod volume estimate (nm ³)	~72.2	~75.2
DAMMIF excluded volume (nm ³)	~92.7	~98.5
Dry volume (nm ³) [from sequence] ‡	~61.4	~64.9
Molecular mass determination		
Calculated monomeric MM (kDa) [from sequence] *	~51	~54
Molecular mass MM (kDa) [from <i>Porod invariant</i>]	45.1 ± 9	47.2 ± 9.4
Molecular mass MM (kDa) [from <i>excluded volume</i>]	46 ± 4.6	49 ± 4.9
Software employed		
Primary data reduction	BRUKER SAS	
Data processing	PRIMUS	
<i>Ab initio</i> analysis	GASBOR	
Validation	DAMAVAR	
Flexibility	EOM	

* http://web.expasy.org/compute_pi/

‡ <http://www.basic.northwestern.edu/biotools/proteincalc.html>

Table 3.6.1 Structural parameters for DENV-4 helicase and hel-NS5₃₂₀₋₃₄₁

The R_g values of DENV-4 helicase determined from Guinier approximation and the $P(r)$ function were 2.50 ± 0.02 nm and 2.41 ± 0.01 nm, respectively, while for DENV-4 hel-NS5₃₂₀₋₃₄₁ were 2.54 ± 0.03 nm and 2.47 ± 0.01 nm, respectively. The D_{max} of the helicase was smaller than hel-NS5₃₂₀₋₃₄₁, which was 7.3 ± 0.7 and 7.6 ± 0.8 nm, respectively. Both R_g and D_{max} values showed the enlargement of particle size in the fused protein hel-NS5₃₂₀₋₃₄₁. The molecular mass determined using Porod invariant and excluded volume derived from the scattering profiles (Table 3.6.1) indicated that both proteins were monomeric at the concentration used. The $P(r)$ functions of the helicase and hel-NS5₃₂₀₋₃₄₁ are shown in figure 3.6.3B, and both exhibited Gaussian distribution which are typical for globular proteins.

Figure 3.6.3 Primary SAXS data analysis of DENV-4 helicase and hel-NS5₃₂₀₋₃₄₁. (A) SAXS scattering patterns of DENV-4 helicase (*black empty circle*) and hel-NS5₃₂₀₋₃₄₁ (*black filled circle*) at concentration of 2.8 mg/ml and 3.0 mg/ml, respectively. The Guinier plots of (left inset) helicase and (right inset) hel-NS5₃₂₀₋₃₄₁ at all three protein concentrations showed linearity, indicating no protein aggregation. (B) The pair-distance ($P(r)$) distribution functions of helicase (*black solid line*) and hel-NS5₃₂₀₋₃₄₁ (*black dashed line*).

3.6.3 *Ab initio* reconstructions of DENV-4 helicase and hel-NS5₃₂₀₋₃₄₁

The program GASBOR was used to restore the *ab initio* envelopes of DENV-4 helicase and hel-NS5₃₂₀₋₃₄₁ using the SAXS profiles shown in figure 3.6.3A. The GASBOR models were reconstructed 10 times independently and the program DAMAVER was used to compute the normalized spatial discrepancy (NSD) for all 10

models. The NSD of 10 models for the helicase was 0.95 ± 0.02 , and for hel-NS5₃₂₀₋₃₄₁ was 0.99 ± 0.03 , indicating that the shape reconstructions of both proteins were reproducible. The least NSD reconstructed model was selected from the helicase (Fig. 3.6.4A) and hel-NS5₃₂₀₋₃₄₁ (Fig. 3.6.4B), and their corresponding fits had the χ^2 of 1.32 and 1.08, respectively.

Figure 3.6.4 *Ab initio* reconstruction of DENV-4 helicase and hel-NS5₃₂₀₋₃₄₁. The superimposition of GASBOR reconstructed solution shapes (*transparent sphere*) of (A) helicase (*green*) and (B) hel-NS5₃₂₀₋₃₄₁ (*cyan*) with the crystallographic structure of DENV-4 NS3₁₇₆₋₆₁₈ (PDB ID: 2JLS), with front (*left*) and side (*right*) views. (C) The overlaid solution shapes of helicase (*green*) and hel-NS5₃₂₀₋₃₄₁ (*cyan*). Protrusion 1 and 2 were indicated as *yellow arrows*. (*inset*) Zoom in view of protrusion 1 region, where the protrusion was close to the proposed NS3-NS5 interacting site. These figures are adapted from Tay *et al.*, 2015 (91).

The crystallographic structure of DENV-4 helicase (NS3₁₇₆₋₆₁₈) (PDB ID: 2JLS (115)) superimposed very well into the helicase *ab initio* model (Fig. 3.6.4A), confirming the reliability of the scattering data and the model reconstruction. After the superimposition of two *ab initio* shapes, two protrusions were identified on the shape of hel-NS5₃₂₀₋₃₄₁, as shown by the yellow arrows in figure 3.6.4C. Protrusion 1 located at the bottom of the shape (Fig. 3.6.4C) led to a more elongated conformation in hel-

NS5₃₂₀₋₃₄₁, which was also indicated by an increase of D_{max} in hel-NS5₃₂₀₋₃₄₁ compared to the helicase. This protrusion was close to the proposed NS3-NS5 interacting site (residues 566-585; *black cartoon* in the *inset* of Fig. 3.6.4C) and the N570 that plays an important role in viral replication (91). The second protrusion, which was located at the side of the helicase, may arise due to conformational changes after the helicase-NS5 peptide interaction. This consistency with data observed from the ELISA experiment (91) suggest a physical interaction between the helicase and NS5 peptide spanning the residues from 320 to 341.

3.6.4 Flexibility assessment of NS5 peptide in hel-NS5₃₂₀₋₃₄₁

To eliminate the possibility that the NS5 peptide in hel-NS5₃₂₀₋₃₄₁ was flexible in solution, the data were further analyzed to assess its flexibility. Similar to lysozyme, the normalized Kratky plots of the helicase and hel-NS5₃₂₀₋₃₄₁ exhibited the bell-shape profiles and had maxima very close to the theoretical maximum of the well-folded globular protein ($\sqrt{3}$, 1.104) (Fig. 3.6.5A). Comparing the curves between two proteins, no change was observed, indicating the introduction of the NS5 peptide did not interfere with the folding of the helicase and both adopted a folded globular conformation.

Figure 3.6.5 Assessment of NS5 peptide flexibility of DENV-4 hel-NS5₃₂₀₋₃₄₁ proteins. (A) The normalized Kratky plots for lysozyme (*grey filled circle*), DENV-4 helicase (*black empty circle*) and hel-NS5₃₂₀₋₃₄₁ (*black filled circle*), and both DENV-4 helicase and hel-NS5₃₂₀₋₃₄₁ showed the bell-shaped profiles often seen in the globular proteins, and overlapped nicely. (B) The R_g distribution of the selected ensembles from DENV-4 hel-NS5₃₂₀₋₃₄₁ exhibited a distribution that has a peak positioned left to the random pool, indicating the helicase and the NS5 peptide are preferably compact in solution.

To assess the compactness of the NS5 peptide with respect to the helicase in hel-NS5₃₂₀₋₃₄₁, the SAXS data of hel-NS5₃₂₀₋₃₄₁ was further analyzed using the program EOM. The R_g distribution of hel-NS5₃₂₀₋₃₄₁ contains a major narrow peak that has a

peak center at 2.46 nm, which was slightly smaller than the maximum of random pool (2.48 nm) (Fig. 3.6.5B), indicating that the helicase and the NS5 peptide are preferably compact and rigid in solution. Altogether, these results demonstrated that the NS5 peptide (residues 320-341) is likely to be bound to the helicase (NS3₁₇₂₋₆₁₈).

3.7 Characterization of DENV-4 NS2B₁₈NS3 in solution using SAXS

NS3 (residues 1–618) is the second largest protein found in DENV. As mentioned earlier, NS3 contributes to the catalytic core in RC together with NS5, and they interact with each other. Similar to NS5, NS3 is a multi-functional protein with two domains, and the domains are linked via a 10/11-residue linker. Its protease domain, which requires the soluble part of NS2B as the cofactor for proper folding, is essential to release the individual proteins from a polyprotein after translation. Its helicase domain is involved in viral RNA replication. The crystallographic structures of DENV-4 NS2B₁₈NS3 (PDB ID: 2VBC, 2WHX; (99,100)) revealed two different domain-domain arrangements, where the protease domain has a 161° rotation with respect to the helicase domain. NS3 from the Murray Valley encephalitis virus (MVES) (PDB ID: 2WV9; (178)) and hepatitis C virus (HCV) (PDB ID: 1CU1; (179)) adopt two different conformations, in which the MVES NS3 has an extended conformation, where the protease domain is spatially separated from the helicase domain, and HCV NS3 has a compact conformation (Fig. 3.7.1). These different conformations suggest that NS3 is a flexible protein. Therefore, in order to get further insight into the flexibility of DENV NS3 in solution, DENV-4 NS2B₁₈NS3 was selected and its flexibility was analyzed using SAXS.

Figure 3.7.1 The domain-domain orientation of NS3 from DENV, MVES and HCV. The protease domain is colored in *slate blue*, the cofactor peptide in *green* and the helicase domain in *orange*. All four crystallographic structures were aligned with each other through the helicase domain. The crystallographic structures of DENV-4 NS3 at conformation I (PDB ID: 2VBC) and conformation II (PDB ID: 2WHX) show the extended conformations, with a 161° rotation in the protease domain with respect to the helicase domain. The MVES NS3 crystallographic structure (PDB ID: 2WV9) has an extended conformation, with the protease domain located slightly away from the helicase. The HCV NS3 crystallographic structure (PDB ID: 1CU1) has a compact conformation.

3.7.1 Expression and purification of DENV-4 NS2B₁₈NS3

The DENV-4 NS2B₁₈NS3 was purified as described in section 2.2.4.2 (Materials and Methods). NS2B₁₈NS3 eluted from the Ni-NTA column at an imidazole concentration of 35 mM to 100 mM, as shown by the SDS gel in figure 3.7.2A. After overnight cleavage with thrombin to remove the thioredoxin-His₆ tag, the sample was subjected to the Ni-NTA column again and NS2B₁₈NS3 without the tag appeared in the flow through (Fig. 3.7.2B). These samples were concentrated and applied to the SEC column. The SEC elution profile showed four peaks, and NS2B₁₈NS3 was identified in the elution fractions corresponding to peak III (*inset* of Fig. 3.7.2C). The elution fractions corresponding to the top 15% of the protein peak (*grey shaded area* in fig. 3.7.2C) were pooled and concentrated to three protein concentrations for SAXS experiment. High purity of the NS2B₁₈NS3 was obtained, as seen in the SDS gel (*inset* of Fig. 3.7.2C).

Figure 3.7.2 Purification of DENV-4 NS2B₁₈NS3. (A) SDS gel after Ni-NTA affinity chromatography for NS2B₁₈NS3. The samples from 50-100 mM imidazole fractions were pooled and subjected to overnight thrombin cleavage to remove the thioredoxin-His₆-tag. (B) SDS gel after overnight cleavage. The cleaved NS2B₁₈NS3 was identified in the flow through (FT). (C) The SEC elution profile of DENV-4 NS2B₁₈NS3. The protein was identified in the elution fractions corresponding to peak III, which was around 15 ml. (*inset*) SDS gel loaded with the fractions from peak III showing high purity of NS2B₁₈NS3.

3.7.2 Primary SAXS data analysis of DENV-4 NS2B₁₈NS3

The SAXS data of DENV-4 NS2B₁₈NS3 were collected immediately after the last step of protein purification at protein concentrations of 1.5, 2.8 and 5.4 mg/ml. The buffer scattering (buffer C: 20 mM Tris/HCl, pH 7.5, 200 mM NaCl, 1 mM DTT) was subtracted from all collected protein data. Three datasets were merged to improve quality of the scattering data, and the merged scattering profile was used for the data analysis (Fig. 3.7.3A). The overall structural parameters derived from SAXS experiments were shown in table 3.7.1. The Guinier plots of all three concentrations showed linearity, indicating good data quality for all the samples and no aggregation were seen in all concentrations (*top inset* of Fig. 3.7.3A). The Kratky plots of DENV-4 NS2B₁₈NS3 showed the protein was well-folded (*bottom inset* of Fig. 3.7.3A). The R_g of DENV-4 NS2B₁₈NS3 derived from the Guinier approximation was determined to be 3.19 ± 0.08 nm, and further confirmed with the R_g derived from the $P(r)$ function, which was 3.23 ± 0.02 nm. The DENV-4 NS2B₁₈NS3 had a $P(r)$ distribution that skewed to the right (Fig. 3.7.3B), indicating an elongated conformation of this protein. The maximum particle dimension D_{max} of DENV-4 NS2B₁₈NS3 protein was determined to be 10.3 ± 1 nm.

Figure 3.7.3 Primary SAXS data analysis of DENV-4 NS2B₁₈NS3. (A) SAXS scattering pattern (*black circle*) of DENV-4 NS2B₁₈NS3 after merging the scattering patterns from the concentrations of 1.5 mg/ml, 2.8 mg/ml and 5.4 mg/ml. (*top inset*) The Guinier plots of all protein concentrations showed linearity, indicating no protein aggregation. (*bottom inset*) The Kratky plot showed that NS2B₁₈NS3 was well-folded. (B) $P(r)$ function of DENV-4 NS2B₁₈NS3.

Protein sample	DENV-4 NS2B ₁₈ NS3
Concentration range (mg ml ⁻¹)	1.50 – 5.40
Structural parameters	
$I(0)$ (arbitrary units) [from P(r)]	78.39 ± 0.64
R_g (nm) [from P(r)]	3.23 ± 0.02
$I(0)$ (arbitrary units) [from Guinier]	78.49 ± 1.13
R_g (nm) [from Guinier]	3.19 ± 0.08
D_{max} (nm)	10.3 ± 1
Porod volume estimate (nm ³)	~97
DAMMIF excluded volume (nm ³)	~131
Dry volume (nm ³) [from sequence] ‡	~86.4
Molecular mass determination	
Calculated monomeric <i>MM</i> (kDa) [from sequence] *	~71.4
Molecular mass <i>MM</i> (kDa) [from <i>Porod invariant</i>]	60.6 ± 12
Molecular mass <i>MM</i> (kDa) [from <i>excluded volume</i>]	65.5 ± 6.6
Software employed	
Primary data reduction	BRUKER SAS
Data processing	PRIMUS
<i>Ab initio</i> analysis	DAMMIF & GASBOR
Validation and averaging	DAMAVAR
Computation of model intensities	CRYSOL
Flexibility	EOM 2.0
* http://web.expasy.org/compute_pi/	
‡ http://www.basic.northwestern.edu/biotools/proteincalc.html	

Table 3.7.1 Structural parameters for DENV-4 NS2B₁₈NS3

3.7.3 Reconstruction of *ab initio* models of DENV-4 NS2B₁₈NS3

Two *ab initio* reconstruction programs, DAMMIF and GASBOR were used to reconstruct the shape of DENV-4 NS2B₁₈NS3 from the scattering data shown in figure 3.7.3A. Twenty independent DAMMIF models were computed, yielding the reproducible models with NSD of 0.774 ± 0.066 . The averaged model of DENV-4 NS2B₁₈NS3 is shown in figure 3.7.4A, and it had an elongated conformation. Ten independent GASBOR models were computed (NSD = 1.282 ± 0.041), and the model with least NSD was shown in figure 3.7.4B. Similar to the DAMMIF model, the

computed GASBOR model revealed an elongated shape, confirming that the shape reconstruction was reliable.

Figure 3.7.4 *Ab initio* modeling of DENV-4 NS2B₁₈NS3. (A) The averaged DAMMIF envelope of DENV-4 NS2B₁₈NS3 from 20 independent reconstructions, with (*left*) front and (*right*) side views. The overall envelope dimensions were 10.5 x 5.4 x 3.5 nm. (B) The GASBOR envelope of DENV-4 NS2B₁₈NS3 with best NSD value from 10 independent reconstructions, with (*left*) front and (*right*) side views.

3.7.4 Structural comparison between the SAXS envelope and crystallographic structure

Up to date, three high resolution structures of DENV-4 NS2B₁₈NS3 have been solved. The first DENV-4 NS2B₁₈NS3 crystallographic structure, 2VBC (100), was superimposed nicely into the DAMMIF envelope using SUPCOMB (Figure 3.7.5A). From these three crystallographic structures, two different conformations were identified, in which the 2VBC structure, which was crystallized from wild-type protein, and the 2WZQ structure (99), which was crystallized from the E173GP174 mutant protein, adopted conformation I. In comparison, 2WHX structure, which was crystallized from the wild-type protein in complex with ADP-Mn²⁺, adopted conformation II that had an approximately 161° rotation of protease domain with respect to the helicase domain. The same conformation II was also identified in the crystal formed from wild-type protein (99). To understand which of these two conformations in solution is predominant, the theoretical scattering curves of these three crystallographic structures were calculated using the program CRY SOL, to evaluate the fit between them and the experimental scattering data of DENV-4 NS2B₁₈NS3. The theoretical experimental curves calculated from 2VBC, 2WZQ and 2WHX fitted nicely into the experimental curve with the discrepancy χ^2 of 0.955, 1.135 and 0.798, respectively. Their fits are shown in figure 3.7.5B. Comparing all

three fittings, 2WHX, which adopted conformation II, had the least discrepancy, suggesting the conformation with ADP-Mn²⁺ bound is slightly more favorable in solution.

Figure 3.7.5 Comparison of SAXS structure and crystallographic structures. (A) Superimposition of the DAMMIF envelope (*grey sphere*) and the crystallographic structure 2VBC (100) (*cartoon representation*) showed the crystallographic structure fits nicely into the envelope. (B) SAXS scattering pattern (black circle) of DENV-4 NS2B₁₈NS3 and calculated scattering profiles of the crystallographic structure 2VBC (*red line*), 2WZQ (99) (*green line*) and 2WHX (99) (*blue line*). All three structures had theoretical profiles that fitted nicely with the experimental scattering pattern, and the crystallographic structure 2WHX, which adopted conformation II, fitted best to the experimental data ($\chi^2 = 0.798$).

3.7.5 Domain mobility of DENV-4 NS2B₁₈NS3

The existence of NS3 crystallographic structures with different domain-domain arrangements suggests movements of both domains respective to each other. To gain some insights into the flexibility of this protein molecule, the data of DENV-4 NS2B₁₈NS3 were further analyzed to obtain this information. Comparing the normalized Kratky plot of NS2B₁₈NS3 with its helicase and reference protein (lysozyme), NS2B₁₈NS3 had a broader bell-curve profile, with a slight shift in its peak maximum (Fig. 3.7.6A). This profile signifies that the NS2B₁₈NS3 has an elongated conformation with less rigidity. This is further confirmed with the Porod-Debye plot of the NS2B₁₈NS3 (Fig. 3.7.6B), where the absence of a Porod plateau and a Porod exponent (P) below 4, indicates that NS2B₁₈NS3 is not compact in solution.

To characterize the flexibility seen in DENV-4 NS2B₁₈NS3, the SAXS data were further analyzed using the program EOM. The selected EOM ensembles generated a solution with a discrepancy χ^2 of 0.225 (Fig. 3.7.7A), and it had a unimodal R_g distribution (Fig. 3.7.7B). NS2B₁₈NS3 exhibited a broad R_g distribution of selected ensembles, and it was numerically compared with the random pool, giving a R_{flex} of

81% compared to the random pool that had the percentage of 87%, with R_σ of 0.81. These indicate that the NS2B₁₈NS3 is flexible in solution, and the protease domain with respect to the helicase adopted a random motion.

Figure 3.7.6 Assessment of protein flexibility in DENV-4 NS2B₁₈NS3. (A) The normalized Kratky plots for lysozyme (*grey filled circle*), DENV-4 NS2B₁₈NS3 (*red filled circle*) and its helicase domain (*black filled circle*). Lysozyme and the helicase showed the typical profile for a compact globular protein, while the NS2B₁₈NS3 had a profile commonly seen in the multi-domain proteins tethered by the linker. (B) The Porod-Debye plots of DENV-4 NS2B₁₈NS3 (*black empty circle*). The data regions used to compute the Porod exponent were shown as the fitted line (*red line*), and its Porod exponent was 3.7, which was smaller from the exponent of 4.0 for a compact globular protein.

Figure 3.7.7 EOM analysis of DENV-4 NS2B₁₈NS3. (A) The ensemble scattering profile (*red line*) calculated for NS2B₁₈NS3 fitted well into the experimental scattering curve (*black empty circle*), with a χ^2 value of 0.225. (B) NS2B₁₈NS3 exhibited a broad unimodal R_g distribution of selected ensembles, indicating NS2B₁₈NS3 is flexible and the two domains adopted a random motion. (*inset*) The arrow indicated the R_g calculated from the

crystallographic structure of NS2B₁₈NS3 (PDB ID: 2VBC). (C) The five representative models for NS2B₁₈NS3 from the EOM calculation with the least χ^2 value. The protease and helicase were represented as *cartoon*, and the helicase was colored in *orange*, while the protease was colored differently. The missing residues, which were re-built, were represented as *sphere*.

3.8 Characterization of DENV-2 capsid protein in solution using SAXS

The capsid (C) protein (residues 1–100) is a highly basic 12 kDa protein. Its structure was obtained in solution using nuclear magnetic resonance (NMR), and revealed that the protein exists as a dimer, with a well-folded domain (residues 20–100) and a conformational labile part (residues 1–20) that was not assigned in NMR structure (33,34). Multiple copies of C protein interact with the viral genomic RNA to form a nucleocapsid. Up to date, the assembly mechanism of C proteins around viral RNA remains unknown. A previous experiment that worked on a small non-enveloped polyomavirus, SV40, showed the kinetic information of the RNA encapsidation by its viral C proteins using time-resolved small-angle X-ray scattering (142). The main aim of this study was to gain insight to the unassigned residues of the NMR structure using SAXS, and the final goal is to obtain the kinetic information of the DENV nucleocapsid formation using SAXS.

3.8.1 Expression and purification of the DENV-2 capsid protein

In the proposed purification protocol described by Jones *et al.*, 2003 (33), the C protein was purified using ammonium sulfate fractionation, followed by cation-exchange chromatography and size-exclusion chromatography. Since the C protein construct prepared in this study contained an N-terminal His-tag, the purification protocol had been modified and optimized. Recombinant C protein was purified using affinity chromatography (Ni-NTA agarose), followed by SEC, using a buffer containing 50 mM Tris/HCl, pH 7.5 and 1 M NaCl. The protein eluted from the Ni-NTA column at an imidazole concentration of 200 mM to 500 mM (Fig. 3.8.1A), and the fractions from 200 mM to 400 mM were pooled together. Protein precipitation was observed while concentrating the sample eluted from the Ni-NTA column and thus, the protein concentrating step was skipped and the eluted sample was applied directly to the SEC column. Two peaks were identified from the SEC elution profile (Fig. 3.8.1B), where the elution fractions corresponding to peak II contained the C protein (*right inset* of Fig. 3.8.1B). RNA (RNase-sensitive) was identified in the elution fractions corresponding to peak I, as shown in the agarose gel (*left inset* of Fig. 3.8.1B). The elution fractions corresponding to the top 15% of the protein peak (*grey shaded area* in Fig. 3.8.1C) were pooled together and concentrated to three protein concentrations for SAXS experiments. As shown in the SDS gel (*right inset* of Fig. 3.8.1B), high protein purity was obtained.

Figure 3.8.1 Purification of DENV-2 capsid proteins. (A) SDS gel after Ni-NTA affinity chromatography. The samples from 200-400 mM imidazole fractions were pooled and applied directly to a SEC column. (B) The SEC elution profile of DENV-2 capsid protein. Two peaks were identified in the elution profile, and the DENV-2 capsid protein was identified in the elution fractions corresponding to peak II. (*left inset*) Nucleic acid agarose gel with fractions from peak I and II. RNase-sensitive nucleic acids were identified in the elution fractions corresponding to peak I. (*right inset*) SDS gel loaded with the fractions from peak I and II. The impurities were identified in the elution fractions corresponding to peak I. The elution fractions of peak II contained the highly purified DENV-2 capsid proteins.

3.8.2 Primary SAXS data analysis of the DENV-2 capsid protein

SAXS data of DENV-2 capsid proteins were collected immediately after the last protein purification step, at protein concentrations of 1.65 mg/ml, 4.05 mg/ml and 7.3 mg/ml. The buffer scattering (buffer D: 50 mM Tris/HCl, pH 7.5, 1 M NaCl, 0.8 mM DTT) was subtracted from all collected protein data. The low angle scattering data of all three protein concentrations (Fig. 3.8.2B) overlapped nicely, indicating no inter-particle interaction between the capsid proteins in all three concentrations. The Guinier regions of all three protein concentrations showed linearity, showing good data quality and no aggregation (*top inset* of Fig. 3.8.2A). Therefore, the data at protein concentration of 4.05 mg/ml, which had less noise, was used for further data analysis. The purified protein was folded, as shown in the Kratky plot (*bottom inset* of Fig. 3.8.2A) exhibited a bell-curve shape with a well-defined maximum. The overall structural parameters derived from SAXS experiments were shown in table 3.8.1. The R_g of DENV-2 capsid protein derived from the Guinier approximation was 2.52 ± 0.02 nm, and confirmed with the R_g derived from its $P(r)$ function, which was 2.63 ± 0.01 nm. The $P(r)$ function of DENV-2 capsid protein showed a main bell curve distribution from 0 to ~5 nm, and a long tail from about 5 nm to 8.5 nm, indicating the particle had a globular structure with an elongated region. The maximum particle dimension D_{max} of DENV-2 capsid protein was determined to be 8.5 ± 0.9 nm.

Protein sample	DENV-2 Capsid
Concentration range (mg ml ⁻¹)	1.65 – 7.3 mg/ml
Structural parameters	
$I(0)$ (arbitrary units) [from $P(r)$]	177.8 ± 0.63
R_g (nm) [from $P(r)$]	2.63 ± 0.01
$I(0)$ (arbitrary units) [from Guinier]	176.7 ± 0.7
R_g (nm) [from Guinier]	2.52 ± 0.02
D_{max} (nm)	8.5 ± 0.9
Porod volume estimate (nm ³)	~42.5
DAMMIF excluded volume (nm ³)	~57.5
Dry volume (nm ³) [from sequence] ‡	~14.2
Molecular mass determination	
Calculated monomeric MM (kDa) [from sequence] *	~11.8
Molecular mass MM (kDa) [from <i>Porod invariant</i>]	26.6 ± 5.3
Molecular mass MM (kDa) [from <i>excluded volume</i>]	28.8 ± 2.9
Software employed	
Primary data reduction	BRUKER SAS
Data processing	PRIMUS
<i>Ab initio</i> analysis	DAMMIF
Validation and averaging	DAMAVR
Computation of model intensities	CRY SOL
Rigid-body modeling	CORAL
Flexibility	EOM 2.0
* http://web.expasy.org/compute_pi/	
‡ http://www.basic.northwestern.edu/biotools/proteincalc.html	

Table 3.8.1 Structural parameters for DENV-2 capsid protein

Figure 3.8.2 Primary SAXS data analysis of DENV-2 capsid protein. (A) SAXS scattering curve (green circle) of the DENV-2 capsid protein at concentration of 4.05 mg/ml. (top inset) The Guinier plots of all three protein concentrations show linearity, indicating no protein aggregation. (bottom inset) The Kratky plot showed that the capsid protein was folded. (B) The SAXS scattering curves at very low angle for all three measured protein concentrations overlapped nicely, indicating no inter-particle interaction between the capsid proteins. (C) $P(r)$ function of the DENV-2 capsid proteins.

3.8.3 Reconstruction of rigid-body models and flexibility analysis of the capsid protein from DENV-2

The molecular masses determined from the Porod invariant and excluded volume from *ab initio* structure (Table 3.8.1) showed that the purified capsid protein is a dimeric protein. Since the atomic resolution of the DENV capsid protein is solved from residues 20 to 100, the theoretical scattering profiles of monomeric or dimeric forms of the NMR structures (PDB ID: 1R6R (34)) were calculated using CRY SOL, and compared with the experimental data. Using the monomeric NMR structure as input, its fit showed a significant difference from the experimental data and had a very large χ^2 value, which is 27.68 (Fig. 3.8.3). However, the dimeric protein fitted better to the experimental data (Fig. 3.8.3), with a discrepancy χ^2 of 8.89, suggesting the capsid protein is a dimeric protein in solution.

Due to the labile conformation of the N-terminal 20 residues of the DENV capsid protein, the first 20 residues are not included in the NMR structure and thus, these residues are missing from the PDB data. Since the capsid protein is in total 100 residues long, the missing residues will contribute significantly to the scattering during

SAXS data collection. This could explain why the fit between dimeric protein and the experimental curve has a significant discrepancy.

Figure 3.8.3 SAXS scattering profile (green circle) of DENV-2 capsid protein, and the calculated scattering profiles of monomeric (red line) or dimeric (blue line) NMR structure 1R6R (34). The dimeric NMR structure fitted better to the experimental scattering data, which had a χ^2 of 8.89, compared to the monomeric NMR structure, which had a χ^2 of 27.68.

Therefore, the rigid-body modeling CORAL was performed using the dimeric NMR structure as the rigid-body and the missing protein chain was re-built. The theoretical scattering profile of the calculated CORAL model fitted nicely into the experimental profile (Fig. 3.8.4A), with a χ^2 of 2.65. The CORAL model rebuilt the missing 20 residues into the NMR structure, and the rebuilt protein stretch was away from the main core of the protein and did not locate to a specific position (Fig. 3.8.4B).

Figure 3.8.4 Rigid-body modeling of DENV-2 capsid proteins. (A) SAXS scattering curve (green circle) of DENV-2 capsid proteins and the fit curve of CORAL model (red line). The calculated rigid-body model has a reasonable good fit to the experimental profile, with a discrepancy of 2.65. (B) The CORAL model of DENV-2 capsid protein in dimeric form, where the *cartoon* represented the NMR structure and the *small spheres* represented the missing 20 residues that were rebuilt from the CORAL.

Since the N-terminus of the capsid protein is un-structured, the SAXS data was further analyzed by considering the protein to be flexible. The normalized Kratky plot was created (Fig. 3.8.5A) from the SAXS data of the capsid protein, showing a bell-shaped profile with a maximum shift away from the maximum of a compact globular protein. This suggests that the capsid protein has a compact region together with a flexible region.

Figure 3.8.5 Assessment of protein flexibility in DENV-2 capsid protein. (A) The normalized Kratky plots for lysozyme (grey filled circle) and DENV-2 capsid protein (green filled circle). Lysozyme showed a typical profile as compact globular protein, while the capsid proteins had a bell-shaped profile with a maximum peak shifted away from the compact globular protein, indicating the presence of flexibility in the capsid protein. (B) The calculated ensemble scattering profile from EOM (red line) fitted well to the experimental scattering profile (black circle), with a discrepancy χ^2 of 2.63. (C) The R_g distribution of selected ensembles from DENV-2 capsid protein exhibited a broad distribution that was positioned right to the random pool, indicating the capsid protein is flexible and has an extended N-terminus. (inset) The R_g calculated from NMR structure of DENV capsid protein (PDB ID: 1R6R, (34)) was located outside the pool. The capsid protein was represented as *cartoon*, and colored in *blue* and *green* for each monomer. (D) The representative models for DENV-2 capsid protein from the EOM calculation with the least χ^2 value.

The flexibility of the capsid protein was further analyzed using EOM. In EOM, the dimeric NMR structure was used as the input, and the missing N-terminal residues were rebuilt by the program. The selected ensemble for DENV-2 capsid protein gave a

discrepancy χ^2 of 2.63 (Fig. 3.8.5B), and the selected representative models are shown in figure 3.8.5D, where the N-terminus of the capsid protein was randomly located in the search volume. The R_g distribution of selected ensembles of the DENV-2 capsid protein had a major distribution between 2.4 nm to 3.1 nm (Fig. 3.8.5C) that was larger than the distribution of random pool, showing that the N-terminus of DENV-2 capsid protein is extended. The R_{flex} , which is a quantitative measurement for flexibility, of DENV-2 capsid protein, was 90%, with a randomness threshold of 86% calculated from the random pool, indicating the capsid protein is flexible, which was also further confirmed by the broad R_g distribution. The quality of the EOM data was confirmed with the R_σ of 1.27, which is expected to be higher than 1.0 when ensemble R_{flex} is more than the pool R_{flex} . Altogether, the EOM analysis of DENV-2 capsid protein indicates that this protein is flexible and has an extended N-terminus.

3.8.4 Salt effect on the DENV-2 capsid protein

In order to understand the mechanism of DENV nucleocapsid assembly, a proper condition for capsid re-assembly reaction is needed. Under 1 M NaCl condition, the capsid protein exists as a stable dimer, as shown in section 3.8.2, indicating the importance of ionic strength to the capsid protein.

The capsid proteins from norovirus were previously shown to associate and dissociate by exchanging the ionic strength in the buffer (143). Since this is the simplest condition, the first trial to obtain the re-assembly reaction condition was performed using the condition in which the ionic strength in the buffer was reduced to its minimum. After overnight dialysis using salt-free buffer, the dialyzed sample was subjected to SAXS data collection, and the data of the dialyzed buffer was collected for background subtraction.

As a control, the SAXS data of the capsid protein from the same purification batch was collected in buffer with 1 M NaCl. The SAXS scattering profiles of DENV-2 capsid protein under two ionic strength conditions are shown in figure 3.8.6A. The overall SAXS structural parameters are shown in table 3.8.2. The linearity seen in the Guinier plots of both conditions indicated no protein aggregation (*inset* of Fig. 3.8.6A). The R_g values derived from the Guinier approximation were 2.59 ± 0.03 nm and 2.35 ± 0.02 nm, for the capsid protein with 1 M NaCl or in the absence of NaCl, respectively. This was confirmed with the R_g values derived from the $P(r)$ functions, where the capsid proteins with 1 M NaCl and in the absence of NaCl had R_g values of

2.63 ± 0.02 nm and 2.44 ± 0.02 nm, respectively. The maximum particle dimension D_{max} of the capsid protein in the presence and absence of 1 M NaCl were determined to be 8.3 ± 0.8 nm and 7.7 ± 0.8 nm, respectively. From the R_g and D_{max} , it is clearly shown that the ionic strength affects the size of the capsid protein. Besides this, as compared to the $P(r)$ function of the capsid protein under 1 M NaCl condition, the $P(r)$ function of the capsid protein in the absence of NaCl had a slight shift toward left, and a shorter tail, signifying that the capsid protein appears to be smaller and less extended (Fig. 3.8.6B). While comparing the normalized Kratky plots of two proteins, they overlapped with each other, indicating no significant change in their flexibility.

Figure 3.8.6 Primary SAXS data analysis of DENV-2 capsid proteins in buffers containing different ionic strength. (A) SAXS scattering curves of DENV-2 capsid proteins in the buffer condition with 1 M NaCl (blue circle) or 0 M NaCl (magenta circle). (inset) The Guinier plots of both conditions showed linearity, indicating no protein aggregation. (B) $P(r)$ functions of DENV-2 capsid proteins in the buffer with 1 M NaCl (blue line) and 0 M NaCl (magenta line). In the absence of NaCl, the capsid protein appeared to be smaller and less extended. (C) The normalized Kratky plots for lysozyme (grey filled circle), and DENV-2 capsid protein in buffer with 1 M NaCl (blue filled circle) or its absence (magenta filled circle). Lysozyme showed a typical profile as a compact globular protein, while the capsid proteins from both conditions had a bell-shaped profile with a maximum peak shifted away from the compact globular protein.

Protein sample	DENV-2 Capsid (1 M NaCl)	DENV-2 Capsid (No salt)
Concentration range (mg ml ⁻¹)	~ 2.3 mg/ml	~ 1.6 mg/ml
Structural parameters		
<i>I</i> (0) (<i>arbitrary units</i>) [from P(r)]	151.2 ± 0.93	199.3 ± 1.01
<i>R_g</i> (nm) [from P(r)]	2.63 ± 0.02	2.44 ± 0.02
<i>I</i> (0) (<i>arbitrary units</i>) [from Guinier]	151.9 ± 1.03	197.0 ± 1.2
<i>R_g</i> (nm) [from Guinier]	2.59 ± 0.03	2.35 ± 0.02
<i>D_{max}</i> (nm)	8.3 ± 0.8	7.7 ± 0.8
Porod volume estimate (nm ³)	~46.0	~35.2
DAMMIF excluded volume (nm ³)	~67.0	~51.0
Dry volume (nm ³) [from sequence] ‡	~14.2	~14.2
Molecular mass determination		
Calculated monomeric <i>MM</i> (kDa) [from sequence] *	~11.8	~11.8
Molecular mass <i>MM</i> (kDa) [from <i>Porod invariant</i>]	28.8 ± 5.8	22.0 ± 4.4
Molecular mass <i>MM</i> (kDa) [from <i>excluded volume</i>]	33.5 ± 3.4	25.5 ± 2.6
Software employed		
Primary data reduction	BRUKER SAS	
Data processing	PRIMUS	
* http://web.expasy.org/compute_pi/		
‡ http://www.basic.northwestern.edu/biotools/proteincalc.html		

Table 3.8.2 Structural parameters for DENV-2 capsid protein in the presence or absence of NaCl

4. DISCUSSION

4.1 Domain-domain arrangement of DENV NS5

DENV NS5 is a two-domain protein, containing the methyltransferase (MTase) and the RNA-dependent RNA polymerase (RdRp) domains. They catalyze two different enzymatic events which are important for virus replication. Together with NS3, which is also a two-domain protein consisting of the protease and helicase domains, they form the catalytic center for the viral genome replication. In addition, NS5 interacts with a variety of host proteins to replicate the viral genome and perturb the host immune response. These inter- and intra-ensemble formation require the enzymes to undergo certain structural alterations and to possess dynamic features.

The high resolution structure of DENV-3 NS5₆₋₈₉₅ had been solved recently using crystallography (PDB ID: 4V0R; (84)). This crystallographic structure reveals a compact structure with its MTase sitting on top of the RdRp finger subdomain, and the active site of the MTase positioned away from the RdRp (Fig. 4.1.1). The crystallographic structure of DENV-3 NS5₆₋₈₉₅ has a maximum length of 8.7 nm (84), and is shorter than the related JEV full-length NS5, which has a maximum length of 10 nm (94). The MTase of JEV NS5 interacts with the RdRp finger subdomain through a hydrophobic interface, and its MTase is rotated by 105° when superimpose the RdRp of both structures, bringing the active site of JEV NS5 MTase closer to the RdRp domain (Fig. 4.1.1).

Figure 4.1.1 Crystallographic structures of (left) DENV-3 NS5₆₋₈₉₅ (PDB ID: 4V0R) (84) and (right) JEV NS5 (PDB ID: 4K6M) (94). Both structures are displayed in *surface representation*. The MTase is colored in *blue*, the 10-residue linker (residues 263 to 272) in *black* and the RdRp is colored according to their subdomains, where the finger subdomain is *red*, the palm subdomain is *green* and the thumb subdomain is *violet*. The bNLS1 protein stretch is colored in *yellow*.

In comparison to NS5 from DENV and JEV, which have the MTase located away from the RNA exit site on the RdRp, WNV NS5 is proposed to have a different domain-domain arrangement. Based on reverse genetic and structural modeling studies, the MTase of WNV NS5 is proposed to locate closer to the RNA exit site of the RdRp (Fig. 4.1.2E). These three different domain-domain arrangements suggest that the MTase and the RdRp domains of the flaviviruses may not adopt a stable conformation. Furthermore, although the two enzymes of NS5 catalyze different enzymatic events, it has been shown that the MTase facilitates the RdRp function by enhancing the template binding affinity (68).

Figure 4.1.2 Possible arrangement of the MTase and the RdRp domains. This figure is taken from Saw *et al.*, 2015 (173). The MTase is colored in *blue*, the linker in *orange* and the RdRp in *magenta*. “F” and “T” denote the finger and the thumb subdomain, respectively. The RNA exit site is indicated by the yellow arrow. (A) The MTase of the crystallographic structure of DENV-3 NS5₆₋₈₉₅ (PDB ID: 4V0R) is positioned on top of the finger subdomain of the RdRp. (B) The MTase of the rigid-body model calculated from the SAXS data is positioned next to the RdRp, with two domains stay closer with each other. (C) The MTase of the highly extended conformation of DENV-3 NS5₁₋₈₇₈ derived from previous SAXS experiment (180) is arranged side-by-side to the RdRp. (D) The MTase of the crystallographic structure of JEV NS5 (PDB ID: 4K6M) is positioned on top of the finger subdomain of the RdRp, with a 105° rotation of the MTase compared to the 4V0R structure. (E) The MTase of the WNV NS5 is proposed to be positioned closer to the RNA exit site (181).

To ascertain the domain-domain arrangement of the DENV NS5 in solution, the solution structure of DENV-3 NS5FL was derived from the SAXS experiment, and it reveals a different arrangement of the MTase and RdRp. Instead of having a compact conformation seen in these crystallographic structures, especially in the DENV-3

NS5₆₋₈₉₅ structure (PDB ID: 4V0R (84)), the solution structures of DENV NS5 adopt a more extended conformation, and the SAXS model suggests that the MTase is positioned next to the RdRp domain (Fig. 3.2.6B). The overall structure has a maximum dimension of 11.5 nm compared to the 8.5 nm in the crystallographic structure. Importantly, the solution structures of NS5FL from the serotypes 1, 2 and 4 also adopt an extended conformation with a similar arrangement of their MTase and RdRp domains (Fig. 3.3.3), having a similar maximum particle dimension ranged from 11.0 to 11.5 nm (Table 3.3.1). To prove that the conformation differences observed from the solution structure and the crystallographic structure are not due to the DENV-3 NS5 construct used in the study, the solution structure of DENV-3 NS5₆₋₈₉₅, which was used in crystallographic study, was also examined, and the same extended conformation was retained. The tight crystal packing arrangement of DENV-3 NS5₆₋₈₉₅ structure (Fig. 3.2.2) may partially explain the difference of the domain-domain arrangement seen in the solution structure and the crystallographic structure of NS5.

4.2 Conformational flexibility of DENV NS5 in solution

Considering the significant difference in the arrangement of the MTase and the RdRp between the solution structure and the crystallographic structure, the flexibility in DENV-3 NS5FL was examined. Assessment of the NS5 flexibility reveals that DENV-3 NS5FL is extended and less rigid in solution (Fig. 3.2.7), and similar outcome is obtained for the crystallographic construct (DENV-3 NS5₆₋₈₉₅). Analyzing the single domains of NS5 in solution reveals that the MTase domain is more flexible and less compact compared to the RdRp domain (Fig. 3.2.10), suggesting the flexibility seen in NS5FL may be contributed by the MTase. The MTase alone catalyzes three major capping activities during viral genome replication and thus, this flexibility may be needed by the MTase to catalyze these sequential steps of the capping mechanism.

The EOM analysis of DENV-3 NS5FL identified an ensemble of conformation that has a unimodal R_g distribution with a main peak around 3.4 to 3.5 nm. The selected models had a set of slight bend and elongated conformations, with the short inter-domain distance between two domains, indicating both domains preferably stay closer to each other. This analysis discloses the flexible nature of NS5 in solution, with the MTase domain arranging into several positions with respect to the RdRp domain. This flexibility may be essential for the functions of this two-domain protein, and also for

the interaction of the NS5 with other viral and host proteins during viral replication. Previous SAXS experiments conducted on DENV-3 NS5₁₋₈₇₈ reported a bimodal R_g distribution with a highly extended conformation that had a dimension of 12.5 nm (180). The MTase and RdRp domains in this highly extended conformation were located apart from each other, as shown in figure 4.1.2C. However, this highly extended conformation could not be identified with the data collected here using freshly purified NS5 protein, as well as for the data of NS5 from serotypes 1, 2 and 4.

4.3 Similarity and diversity of the NS5FLs from DENV-1 to -4

Previous studies showed that the nuclear localization of NS5s is different among DENV-1 to -4 (73,74). Despite high sequence similarity across the four serotypes, the linker that tethers the two enzymatic domains have low sequence similarity. To gain structural insights into the similarity or the difference of the NS5FLs from DENV-1 to -4, solution data of these four proteins were collected and compared. The solution structures of NS5 from all four serotypes have a similar overall structure, with an extended conformation (Fig. 3.3.3) and a maximum dimension ranging from 11.0 to 11.5 nm (Table 3.3.1). The MTase domain of their SAXS models is positioned next to the RdRp domain. The EOM data reveal a unimodal distribution for the NS5FLs from all four serotypes (Fig. 3.3.5). These similarities confirm that the NS5 from DENV is extended in solution.

Despite these similarities, the solution data of these four NS5FLs reveal a flexibility difference in the DENV-4 NS5FL compared to other three DENV serotypes. The R_g values derived from DENV-4 NS5FL data are significantly smaller compared to NS5FLs of DENV-1, -2 and -3 (Table 3.3.1). Comparison of their normalized Kratky plots and $P(r)$ functions showed a decrease in flexibility and extension of DENV-4 NS5FL (Fig. 3.3.4). The EOM data also showed that the major peak of R_g distribution of DENV-4 NS5FL is shifted to a smaller R_g value. These data reflect that DENV-4 NS5FL has a more compact conformation and difference in its flexibility relative to the NS5FLs from DENV-1 to -3. The increase in the extension and the flexibility seen in the normalized Kratky plots and Porod-Debye plots of the full-length NS5 as well as in the individual domains of NS5 that contain the linker residues, suggest that the 10-residues linker that tethers the two domains may be a good candidate in providing this flexibility. Therefore, to disclose the role of linker in this phenomenon, the 10-residue linker from DENV-3 NS5FL was swapped into DENV-4 NS5FL, and *vice*

versa. Compared to wild-type DENV-3 NS5FL, the substitution of DENV-4 NS5FL linker into DENV-3 NS5FL did not trigger any significant change in its flexibility, while the substitution of DENV-3 NS5FL linker into DENV-4 NS5FL resulted in a smaller D_{max} and a greater compactness compared to wild-type DENV-4 NS5FL. These results imply the importance of the linker residues of DENV-4 NS5 in maintaining the flexibility of the protein. Besides this, mutagenesis studies in the linker region of DENV-4 NS5FL, by substituting R263 and E270 to an alanine, and V264 to other residue led to the inhibition of the MTase and RdRp activities and/or failed to detect any viral RNA and viable viruses (84,95), demonstrating the importance of the linker in the NS5 protein function.

Since the linker of DENV-4 NS5 plays a role in its flexibility and protein functions, this led to the next postulation that the difference observed may be due to the different amino acid partners in the linker region and the MTase/RdRp domains of the D4NS5-d3linker, as the NS5 linkers have a very low sequence conservation (Fig. 3.1.1) among the four serotypes. To pinpoint the residue(s) that are responsible for this event, the non-conserved amino acids in the DENV-4 NS5 linker were substituted with the amino acids from the DENV-3 NS5 linker. This mutational SAXS study identified three key residues, S266, T267 and K271, from the linker of DENV-4 NS5, are needed to keep the proper compactness and flexibility in the DENV-4 NS5FL. Mutations of these three residues led to the enlargement and increase of flexibility compared to wild-type DENV-4 NS5FL. Their $P(r)$ functions and normalized Kratky plots (Fig. 3.4.5), as well as their major ensemble populations selected from EOM analysis (Fig. 3.4.6), are similar to the wild-type DENV-3 NS5FL, instead of wild-type DENV-4 NS5FL.

Looking into the crystallographic structure of DENV-3 NS5₆₋₈₉₅ (PDB ID: 4V0R), the residues S266 and T267 from DENV-4 NS5FL, which correspond to the N265 and A266 of DENV-3 NS5FL, respectively, are located at the center of the 3_{10} helix (*red arrow* in Fig. 4.3.1), which is proposed to function as a swivel in the structure (84). Structural perturbation may arise when these two residues are deleted or substituted with other residues, leading to an alteration in the conformation and/or flexibility of the DENV-4 NS5FL. In addition, these two residues are polar amino acids and exposed to solvent and thus, they may form additional interactions with the neighboring residues which in turn, reduce the flexibility in the two domains.

Figure 4.3.1 Interaction of the NS5 linker residues (*black stick*) and the MTase (*blue stick*) or the RdRp (*red stick*) analyzed from the 4V0R structure (84). The figure is adapted from Subramanian Manimekalai *et al.*, 2016 (176). H263 forms the hydrogen bonding with T261 of the MTase. The linker region from residues V264 to E267 forms a 3_{10} helix, and the conserved residue E267 forms a hydrogen bonding or salt bridge with the residues of both the MTase and the RdRp domains. E269 and T270 interact with K595 of the RdRp. N272 forms the hydrogen bonding with V275, I276 and N574 of the RdRp.

Comparing the linker between DENV-3 and DENV-4 NS5, the linker of DENV-4 NS5 is more polar than the NS5 of DENV-3. The non-polar residues A266 and P268 of DENV-3 NS5 are substituted with the polar residues T267 and T269 in DENV-4, and uncharged residues T270 and N272 are substituted with the charged residues K271 and D273. The changes in the property of linker residues may change the interacting pair between linker and the NS5 domains, thereby changing the flexibility in the DENV-4 NS5. Furthermore, DENV-4 NS5 has a lysine residue with a longer side chain at position 271, and mutation of this residue to a shorter side chain T270 (*brown arrow* in Fig. 4.3.1) may eliminate the interaction between this residue and other residue, which in turn may reduce the flexibility of the two-domain protein. Interestingly, with the double, triple and quadruple mutations involving this residue, the compactness of the mutants are recovered and adopted similar features as the wild-type DENV-4 NS5FL (Fig. 3.4.5). As the linker of NS5 is exposed to the solvent, mutating two or more linker residues from DENV-4 NS5FL to DENV-3 NS5FL make the linker region becomes less polar, which in turn may compensate the flexibility provided by the K271T mutant and restrict the flexibility in the NS5.

4.4 Substrate binding of NS5

NS5 comprises two active enzymes that require different substrates for their catalytic activities. The functions of the MTase domain in the capping mechanism include the transfer of GMP group to the 5' end of RNA, the transfer of the methyl group to the N-7 position of guanine and to the ribose-2'-O position of the first nucleotide of the RNA. To catalyze these reactions, the MTase has to interact with RNA, GTP and SAM. The RdRp domain is responsible for the synthesis of a new strand RNA, and this requires RNA-binding. During the catalysis, the enzyme is proposed to undergo conformational changes in order to bind, catalyze and release the substrate or product. Studying these events can provide insights into the catalytic mechanism including the transition states.

Using the synchrotron radiation provided by the SSRL, SAXS data of DENV-2 NS5FL with GTP, GDP, SAM, SAH and RNA were recorded for the first time. The static SAXS experiments were performed in order to obtain the initial and final states of NS5 upon substrate binding. Although no major change in the overall conformation was observed after adding GTP, GDP, SAM and SAH to protein solution, a repeatable increase in the R_g -values was detected for NS5 in the presence of Gppp-RNA and cap-0 RNA (Table 3.5.1). Further analysis reveals a slight conformation and flexibility changes in NS5 upon binding to both RNAs (Fig. 3.5.2). Same trends were also observed in another set of experiment conducted with DENV-4 NS5FL and phosphorylated RNA (phosRNA) (Fig. 3.5.3). It confirms that the binding of RNA induces a change in NS5.

To shed light on the possible intermediate states, time-dependent SAXS experiments were performed. Similar to the static SAXS experiment, a significant size change was observed in NS5 after interaction with RNA (Fig. 3.5.4 and 3.5.5), regardless of protein concentrations and the protein-to-substrate ratio. However, the change was seen within one minute after mixing NS5 with RNA, indicating that a fast time resolution maybe needed. Therefore, time-resolved SAXS experiments were performed at a time resolution of 20 ms. Upon adding RNA, the R_g of NS5 increased (Fig. 3.5.6C), but due to the high standard deviation for each time point, no conclusive interpretation can be provided. Therefore, to analyze the interaction between NS5 and RNA in this time resolution, the data collection of this sample has to be repeated several times to minimize this standard deviation in the future.

4.5 NS3-NS5 interacting epitope

The flavivirus replication complex is composed of at least of two multifunctional proteins, NS5 and NS3, that play a central role in the viral genome replication. These two proteins interact with each other, and their enzymatic activities might be interdependent in the RC. Previous studies mapped the interacting sites between both proteins to the bNLS-region of the NS5 RdRp (residues 320-368) and the subdomain III of NS3 helicase (residues 566-585) based on biochemical, genetic and biophysical approaches (71,90,91). However, no structural information of this NS3-NS5 complex was described so far. Therefore, SAXS was employed in this study to obtain the structural information of the interaction between the NS3 helicase (residues 172-618) and the bNLS-region of NS5 (residues 320-341).

The comparison of the solution shapes between the helicase (NS3₁₇₂₋₆₁₈) and hel-NS5₃₂₀₋₃₄₁ (NS3₁₇₂₋₆₁₈-NS5₃₂₀₋₃₄₁) suggests a physical interaction between the helicase and the NS5 peptide (NS5₃₂₀₋₃₄₁) (Fig. 3.6.4C). The protrusion observed in the hel-NS5₃₂₀₋₃₄₁ is close to the proposed interacting site of NS3, residues 566-585, and also to the critical interacting residue N570 (*black cartoon representation in the inset of Fig. 3.6.4C*) (91). In addition, this interaction may induce a slight change in the helicase, giving rise to the second protrusion in the solution shape of hel-NS5₃₂₀₋₃₄₁ (Fig. 3.6.4C). To ensure that the NS5 peptide is not moving randomly in solution, EOM analysis was performed, and the generated R_g distribution confirmed that the hel-NS5₃₂₀₋₃₄₁ is preferably compact in solution.

Figure 4.5.1 Schematic diagram of NS3-NS5 interaction. This figure is modified from Tay *et al.*, 2015 (91). NS3 and NS5 are colored according to their subdomains, where in NS3, *yellow*, *orange* and *red* represented subdomain I, II and III, respectively, and in NS5, *red*, *green* and *violet* represented fingers, palm and thumb subdomains. The *blue surface* indicated the interacting peptides from NS3 and NS5.

The SAXS results are further confirmed *in vitro* and *in vivo* in the same studies (91). In the competitive ELISA experiments, the peptide NS3₅₆₆₋₅₈₅ interrupted the NS3-NS5 interaction, and *in vivo*, both NS5₃₂₀₋₃₄₁ and NS3₅₆₆₋₅₈₅ peptides reduced the viral replication. The highly conserved residue N570 in NS3 (NS3:N570A) was shown to be important to the NS3-NS5 interaction *in vitro*, and mutation of this residue could abolish the virus production (91), which is similar to the mutation of K330 (NS5:K330A) in NS5 (96). Unlike the NS5:K330A mutant, which has the mutation that affects RNA synthesis, the NS3:N570A mutant could synthesize very low amount of viral RNA and proteins, even though it is unsustainable to the virus. Compared to the WT that showed synchronized synthesis of (+) and (-) stranded RNAs, NS3:N570A mutant showed an accumulation of (-) stranded RNA as a double-stranded RNA form. This observation indicates the importance of NS3-NS5 interaction in modulating RNA replication in the virus.

4.6 Structural characterization of NS3 from DENV-4 in solution

The interaction between NS3 and NS5 was shown to be important for the DENV replication. A mutational study of the critical interacting residue N570 of NS3 abolished virus production and affected the balanced synthesis of positive- and negative-sense RNA (91). Being an important interacting partner of NS5 and a two-domain protein, solution studies were carried out for DENV-4 NS2B₁₈NS3, to assess its structural traits and flexibility in solution. Comparing the solution shape and crystallographic structure (PDB ID: 2VBC (100)) of DENV-4 NS2B₁₈NS3, both structures showed an extended conformation, and the crystallographic structure (2VBC) superimposed nicely into the solution structure (Fig. 3.7.5A).

From the same protein construct, two domain-domain orientations were reported based on the crystallographic structures, termed as conformation I (PDB ID: 2VBC (100) and 2WZQ (99)) and conformation II (PDB ID: 2WHX (99)). By aligning both conformations centered at the helicase domain, the structural comparison reveals that the protease domain in conformation II has a 161° rotation compared to conformation I. This rotation does not change the extended conformation seen in the crystallographic structure (2VBC), but the ATP-binding site is now exposed to the solvent, as shown as the *light green surface* in figure 4.6.1. In conformation I, the ATP-binding site is blocked by the protease domain, preventing the entry of ATP into the helicase.

However, in conformation II, the protease domain is rotated away, and the ATP-binding site is exposed, allowing ADP-Mn²⁺ to be bound to the helicase domain.

Figure 4.6.1 Comparison of domain-domain orientations of DENV-4 NS2B₁₈NS3 in conformation I (PDB ID: 2VBC (100)) and II (PDB ID: 2WHX (99)). (Left) Superposition of the 2VBC (red cartoon) and 2WHX (blue cartoon) structures. (Middle) The protease domain (light red) of the 2VBC structure (surface representation) blocks the ATP-binding site (light green) located at the helicase domain (red). (Right) The protease domain (blue) of the 2WHX structure rotates 161° away, exposing the ATP-binding site (light green) and enabling binding of ADP-Mn²⁺ (stick representation) to the helicase domain.

Interestingly, the solution data collected for DENV-4 NS2B₁₈NS3 showed that conformation II is slightly more favorable in solution (Fig. 3.7.5B). The protease domain of NS2B₁₈NS3 adopts a random motion with respect to the helicase domain in solution. This flexibility is also observed in both crystallographic structures, where the reported B-values in the protease domain are higher, and it is also consistent with proteolysis data, demonstrating that the protease domain is more labile (99). Compared to the crystallographic structure (2VBC), which has a R_g -value of 2.98 nm, the major population of the selected ensemble lies between 3.2 nm to 3.3 nm in the R_g distribution. This indicates that the inter-domain distance of the two domains of NS2B₁₈NS3 is predominantly extended in solution. The linker region of NS2B₁₈NS3 is suggested to provide some intrinsic flexibility to the protein because of its significant higher B-value in the conformation II (99). Knowing that the protease domain of this protein cleaves in *cis* at the NS2A-NS2B junction and in *trans* at the NS3-NS4A junction (182,183), it is not surprising that the protease domain with respect to the helicase domain adopts a flexible motion, so that the protease can re-orientate and extend itself, to carry out its proteolytic function.

Furthermore, as shown by the crystallographic structures of both conformations, the position of the protease with respect to the helicase can shield or expose the ATP

binding site. It was previously shown that the ATPase activity of the helicase was reduced about 70% in the absence of the protease domain (99). In addition, substitution of the amino acids in the linker region caused a significant drop in ATPase- and helicase activities, as well as RNA replication (99). Therefore, to carry out both the protease and the helicase activities optimally, the inter-domain linker of NS3 can provide this flexibility to the NS2B₁₈NS3.

In the SAXS studies of NS5 presented, DENV-4 NS5 revealed a more compact and less flexible form compared to other three serotypes. This difference in flexibility may be compensated by the flexibility contributed from its interacting partner NS3. To gain further insight into this issue, a detailed analysis of NS3 of DENV-1 to -3 may be required. Besides this, the NS3-NS5 complex study using SAXS can provide more insights to this replication assembly in solution, to understand the molecular machinery required for the viral RNA replication.

4.7 Structural characterization of the capsid protein from DENV-2 in solution

The NMR structure of DENV-2 capsid protein (PDB ID: 1R6R (34)) reveals a dimer structure that contains a well-folded domain (residues 20-100) and a conformational labile N-terminal region (residues 1-20), which is not included in the reported NMR structure. The structure showed an asymmetrical charge distribution, where the N-terminal residues located at the top side of the dimer are hydrophobic, and the C-terminal residues at the bottom are mainly positively-charged (Fig. 4.7.1). The hydrophobic groove is formed by the residues 45 to 65, reported to be involved in membrane integration (184). The positively-charged amino acids at the C-terminus provide a surface for the interaction of the negatively-charged RNA.

Due to the conformational flexibility of the first 20 residues of the capsid protein, these residues are not included in the NMR structure. To gain some structural information of this N-terminal region, the full-length capsid protein from DENV-2 was analyzed in solution using SAXS. This study confirmed the dimerization of the capsid protein, as indicated by the derived molecular mass from the scattering data, and the CRY SOL fitting (Fig. 3.8.3). Using the scattering data of the full-length capsid protein, the missing 20 residues were shown to form a random chain (Fig. 3.8.4B). The ensemble analysis showed that this protein stretch is highly flexible, and extended out from the core globular part of the capsid protein.

Figure 4.7.1 The NMR structure of DENV-2 capsid (residues 20-100) (PDB ID: 1R6R; (34)) represented as *cartoon* (*top*) and *charge distribution surface* (*bottom*). The charge distribution surface is colored in *blue* for positively-charged residues, *white* for uncharged and *red* for negative-charged amino acid. Both representations are showed in top (*left*), side (*middle*) and bottom (*right*) view. The top surface is preferentially distributed by the hydrophobic residues (*red arrow* in top view), while the bottom surface is highly positively-charged (*red arrow* in bottom view).

Interestingly, the particle size of the DENV-2 capsid protein becomes smaller when the buffer's ionic strength is decreased. Looking into the amino acid sequence of the protein, the flexible N-terminal region is charged. By reducing the ionic strength from the protein solution, this N-terminal region may start to fold toward the globular core of the capsid protein, or form certain secondary structures as seen in the capsid peptide 14-23 that forms an α -helix in the presence of phospholipid (127) and thus, reducing the size of capsid protein.

As shown in this study, the ionic strength has effect in the conformation of the DENV-2 capsid protein, but it does not trigger the self-assembly of the capsid. Previous study showed that the *in vitro* assembly of nucleocapsid-like particle can be achieved by mixing the purified capsid protein with a random oligonucleotide, regardless of DNA or RNA (123). In future, this study will be further extended by performing the re-assembly of capsid under this condition. Once the re-assembly condition is established, the re-assembly event will be studied using time-resolved SAXS experiment, to obtain the kinetic information of the DENV nucleocapsid formation.

4.8 Overview of SAXS technique

Structural biology studies the molecular structure of the biological macromolecules, providing insights into their biological function in terms of structures. Several techniques can provide structural information, including macromolecular crystallography, nuclear magnetic resonance (NMR), electron microscopy (EM) or SAXS. Each of these techniques comes with its advantages and weaknesses. For example, macromolecular crystallography provides an atomic resolution structure, which allows the structural study at atomic level, but this technique requires crystals that are uniform and of good quality. Hence, the chances of crystal formation are reduced when dealing with a flexible protein. NMR characterizes structures at atomic resolution and studies the macromolecule dynamics in solution, but this technique is limited by the molecular mass (<70 kDa). To compensate the problems faced in these techniques, EM and SAXS can be used, especially to the larger macromolecular assemblies, to provide more understanding about macro- or supramolecular complexes, as well as their conformational changes while performing their biological functions.

SAXS is a solution technique that offers structural and dynamic information for small to large macromolecules covering a range of sizes that are limited in NMR and EM methods (130). Even though SAXS is limited by the detailed atomic information, it enables the determination of the low resolution structure of macromolecules in solution without any further experimental step, which may be time consuming or could possibly introduce unfavorable stress to the macromolecule. Moreover, SAXS allows the structural study on a flexible system, and also on systems involved in large assembly formation, including protein-protein complexes or protein-substrate formation. Therefore, SAXS is a complementary tool to other structural techniques.

In this thesis, SAXS has been shown to provide a good comparison of a multi-domain protein in the crystal lattice and in solution. Comparing the structures of the DENV-3 NS5 in the crystal lattice (Fig. 3.2.1) and in solution (Fig. 3.2.5), it is clearly shown that in solution, NS5 adopts a more extended conformation, which provides more surface area for the interaction with different viral and host proteins, as well as its substrates. A clearer difference can be seen from the comparison of the theoretical scattering curve with the experimental scattering curve, yielding a high discrepancy among them (Fig. 3.2.6A). The compact conformation seen in the crystallographic structure may be due to the crystal lattice force (Fig. 3.2.2). Compared to NS5, the

solution structure of DENV-4 NS2B₁₈NS3 is in agreement with the crystallographic structure, confirming the overall arrangement of their domains (Fig. 3.7.5). The theoretical scattering profiles of the crystallographic structures calculated using CRY SOL fit nicely into the experimental data, proving the domain-domain arrangement of this protein in solution. Therefore, on top of the detailed atomic information provided by the crystallographic structure, SAXS can add structural details such as the domain-domain arrangement and provide a validation of the crystallographic structure in solution.

To increase the chance of a protein crystal formation, the flexible region of the protein, especially the N- or the C-termini, are often removed, to achieve conformational homogeneity. Besides this, the loops or some parts of the protein that possess inherent flexibility and/or conformational heterogeneity are often invisible in crystallography and NMR. Using the data collected from SAXS, modeling using dummy residues (DRs) can re-build these missing residues into the incomplete high resolution structure. In this thesis, SAXS provides the information required to rebuild the missing fragments of the proteins using the rigid-body modeling software BUNCH and CORAL, and the flexibility analysis software EOM. The missing fragments in the RdRp domain of the NS5 (residues 407-419 and 455-469) were rebuilt using EOM and they are shown to bury inside the RdRp. Since the two domains of NS5 in solution are orientated different when compared to the crystallographic structure, the linker was removed during SAXS modeling, and rebuilt back in the final model. In the case of DENV-2 capsid protein, due to the conformational labile of the first 20 residues in this protein, they are not included in the NMR structure. To gain an idea about the spatial arrangement of these N-terminal residues, they had been rebuilt into the main core protein using collected SAXS data, and this protein stretch adopted a random arrangement in solution (Fig. 3.8.4).

SAXS is capable of handling a polydispersed system where the sample solution contains different types or populations of particles. This allows SAXS to be used to analyze the flexibility and conformational variety of the proteins. In this thesis, the flexibility of two multi-domain proteins, NS3 and NS5, were assessed using SAXS. These two proteins contain two domains tethered by a linker. By transforming the scattering data to the normalized Kratky plot and the Porod-Debye plot, these two proteins appear to be flexible in solution. Using an ensemble approach, EOM, an

ensemble of 5-6 structures was selected for representation, and the two domains were arranged spatially different in solution. Interestingly, in this thesis, SAXS was shown to be capable to detect the compactness difference among the NS5s from four different DENV serotypes. This difference was further assessed by performing mutagenesis studies on the linker region of NS5, and together with SAXS, the critical linker residues being responsible for the NS5 flexibility were identified.

Furthermore, SAXS can be used to track down the conformational change of proteins upon substrate binding. The advancement on the synchrotron radiation, as well as the fast readout detector allows a good quality dataset with low error to be collected. In this thesis, the substrate binding study was performed on the NS5 at SSRL, USA. No significant change was observed in the NS5 conformation for the small molecule binding, but a repeatable change in size was observed when NS5 interacts with RNA. To understand this phenomenon and also the transition states that may appear during NS5-RNA interaction, the substrate binding study was further carried out in a time-dependent manner. In this study, time-dependent SAXS experiments showed that NS5 underwent an overall size change upon RNA-binding. In order to achieve a deeper insight into the fast catalytic events, time-resolved SAXS (TR-SAXS) experiments in a 20 ms time scale were performed, but no major overall change of NS5 could be observed. However, due to the large standard deviation, these experiments have to be repeated in order to obtain a significant statistic. Besides NS5, the capsid protein from DENV-2 will be a good candidate for future TR-SAXS experiments, with the aim of getting the kinetic information of its assembly.

5. CONCLUSION

In this thesis, SAXS experiments were performed to study the NS3, NS5 and capsid protein of DENV. With the existing structural information provided by the X-ray crystallography and NMR, SAXS extends the structural knowledge of these proteins, and provides dynamic information of these proteins. DENV NS5 and NS3 are two important proteins that form the catalytic core for viral genome replication. As compared to the previously reported crystallographic structure of DENV-3 NS5₆₋₈₉₅ (PDB ID: 4V0R) (84), this thesis presented a different domain-domain arrangement of the NS5 that suggests structural alteration and transition of NS5 in solution, which may be crucial for facilitating the interaction of NS5 with a variety of viral and host proteins. The MTase was shown to be the main contributor for this flexibility, which could be essential for the MTase to carry out its three different enzymatic events. DENV-4 NS5 adopted a slightly more compact conformation than the other three serotypes. The mutagenesis studies on the DENV NS5 linker demonstrate the importance of these linker residues in maintaining the proper compaction in NS5, as well as its flexibility. Future studies could focus on the *in vivo* studies of different virus mutants to understand the functional roles of differences in their structures and flexibility. While no significant conformational change in NS5 was observed upon adding different substrates, the addition of RNA to NS5 altered the particle size of the protein, signifying a change in conformation of NS5 upon RNA binding. To get further insight into this phenomenon, the interaction of RNA with the individual domains can be characterized in the future, as well as the binding kinetics of all three protein-RNA pairs, through biophysical methods such as isothermal titration calorimetry (ITC).

In this thesis, the first structural information to the interaction between the NS3 and NS5 was assessed through SAXS experiment. Apart from this, the solution data of DENV-4 NS2B₁₈NS3 revealed a favorable conformation in solution, which is a confirmation that allows the binding of ADP-Mn²⁺ into the nucleotide-binding pocket of helicase domain. As compared with its partner NS5, this protein exhibits a more flexible conformation, which may be important for the protease domain to carry out its proteolytic function. Future studies could be carried out by performing a mutagenesis study on the linker region that prevents the random movement of the protease, and assess its impacts in both the structural and functional roles. Besides the non-structural proteins, the capsid protein, which is the structural protein that encapsidates RNA to

form the nucleocapsid core, was analyzed in solution using SAXS. The SAXS data of DENV-2 full-length capsid protein showed that the first 20 missing residues in NMR structure form a random chain that is highly flexible and extended in solution. As shown in this thesis, unlike in the case of norovirus capsid protein, DENV capsid protein cannot undergo self-assembly by monitoring the buffer ionic strength. However, the particle size of this protein was changed upon changing the ionic strength in the solution which was hypothesized here, be due to the folding of the N-terminal region toward the globular core of the capsid protein. To obtain an ideal self-assembly condition for the study of self-assembly kinetics of capsid protein, more conditions such as changing pH or adding RNA have to be tested in future.

6. REFERENCES

1. International Committee on Taxonomy of Viruses (ICTV). Virus taxonomy: 2014 release.
2. Lindenbach, B. D., Thiel, H.-J., and Rice, C. M. (2007) *Flaviviridae: The viruses and their replication*. in *Fields virology* (Knipe, D. M., and Howley, P. M. eds.), 5 Ed., Lippincott-Raven Publishers, Philadelphia. pp 1101 - 1152
3. Roehrig, J. T. (2003) Antigenic structure of flavivirus proteins. in *Adv. Virus Res.*, Academic Press. pp 141-175
4. Calisher, C. H., and Gould, E. A. (2003) Taxonomy of the virus family *Flaviviridae*. in *Adv. Virus Res.*, Academic Press. pp 1-19
5. World Health Organization (WHO). Impact of Dengue.
6. World Health Organization (WHO). (2011) Comprehensive guidelines for prevention and control of Dengue and Dengue haemorrhagic fever. Revised and expanded Ed., World Health Organization (WHO), Regional Office for South-East Asia, New Delhi, India
7. Kyle, J. L., and Harris, E. (2008) Global spread and persistence of Dengue. *Annu. Rev. Microbiol.* **62**, 71-92
8. Halstead, S., and O'Rourke, E. (1977) Dengue viruses and mononuclear phagocytes. I. Infection enhancement by non-neutralizing antibody. *J. Exp. Med.* **146**, 201-217
9. Halstead, S. B., and O'Rourke, E. J. (1977) Antibody-enhanced Dengue virus infection in primate leukocytes. *Nature* **265**, 739-741
10. Rodenhuis-Zybert, I. A., Wilschut, J., and Smit, J. M. (2010) Dengue virus life cycle: viral and host factors modulating infectivity. *Cell. Mol. Life Sci.* **67**, 2773-2786
11. Halstead, S. B. (2007) Dengue. *The Lancet* **370**, 1644-1652
12. Wagner, E. K. (2008) *Basic virology*, 3 ed., Malden, MA ; Oxford : Blackwell Pub., 2008
13. Schmid, M. A., Diamond, M. S., and Harris, E. (2014) Dendritic cells in Dengue virus infection: Targets of virus replication and mediators of immunity. *Front. Immunol.* **5**
14. Wu, S.-J. L., Grouard-Vogel, G., Sun, W., Mascola, J. R., Brachtel, E., Putvatana, R., Louder, M. K., Filgueira, L., Marovich, M. A., Wong, H. K., Blauvelt, A., Murphy, G. S., Robb, M. L., Innes, B. L., Birx, D. L., Hayes, C. G., and Frankel, S. S. (2000) Human skin Langerhans cells are targets of Dengue virus infection. *Nat. Med.* **6**, 816-820
15. Jessie, K., Fong, M. Y., Devi, S., Lam, S. K., and Wong, K. T. (2004) Localization of Dengue virus in naturally infected human tissues, by immunohistochemistry and *in situ* hybridization. *J. Infect. Dis.* **189**, 1411-1418

16. Cruz-Oliveira, C., Freire, J. M., Conceição, T. M., Higa, L. M., Castanho, M. A. R. B., and Da Poian, A. T. (2015) Receptors and routes of Dengue virus entry into the host cells. *FEMS Microbiol. Rev.* **39**, 155-170
17. McGreal, E. P., Miller, J. L., and Gordon, S. (2005) Ligand recognition by antigen-presenting cell C-type lectin receptors. *Curr. Opin. Immunol.* **17**, 18-24
18. Pokidysheva, E., Zhang, Y., Battisti, A. J., Bator-Kelly, C. M., Chipman, P. R., Xiao, C., Gregorio, G. G., Hendrickson, W. A., Kuhn, R. J., and Rossmann, M. G. (2006) Cryo-EM reconstruction of Dengue virus in complex with the carbohydrate recognition domain of DC-SIGN. *Cell* **124**, 485-493
19. Tassaneetrithep, B., Burgess, T. H., Granelli-Piperno, A., Trumpfheller, C., Finke, J., Sun, W., Eller, M. A., Pattanapanyasat, K., Sarasombath, S., Birx, D. L., Steinman, R. M., Schlesinger, S., and Marovich, M. A. (2003) DC-SIGN (CD209) mediates Dengue virus infection of human dendritic cells. *J. Exp. Med.* **197**, 823-829
20. Lozach, P.-Y., Burleigh, L., Staropoli, I., Navarro-Sanchez, E., Harriague, J., Virelizier, J.-L., Rey, F. A., Desprès, P., Arenzana-Seisdedos, F., and Amara, A. (2005) Dendritic cell-specific intercellular adhesion molecule 3-grabbing non-integrin (DC-SIGN)-mediated enhancement of Dengue virus infection is independent of DC-SIGN internalization signals. *J. Biol. Chem.* **280**, 23698-23708
21. Littaua, R., Kurane, I., and Ennis, F. A. (1990) Human IgG Fc receptor II mediates antibody-dependent enhancement of Dengue virus infection. *J. Immunol.* **144**, 3183-3186
22. Rodrigo, W. W. S. I., Block, O. K. T., Lane, C., Sukupolvi-Petty, S., Goncalvez, A. P., Johnson, S., Diamond, M. S., Lai, C.-J., Rose, R. C., Jin, X., and Schlesinger, J. J. (2009) Dengue virus neutralization is modulated by IgG antibody subclass and Fc γ receptor subtype. *Virology* **394**, 175-182
23. Heinz, F. X., and Allison, S. L. (2000) Structures and mechanisms in flavivirus fusion. in *Adv. Virus Res.*, Academic Press. pp 231-269
24. Mukhopadhyay, S., Kuhn, R. J., and Rossmann, M. G. (2005) A structural perspective of the flavivirus life cycle. *Nat Rev Micro* **3**, 13-22
25. Perera, R., and Kuhn, R. J. (2008) Structural proteomics of Dengue virus. *Curr. Opin. Microbiol.* **11**, 369-377
26. Junjhon, J., Pennington, J. G., Edwards, T. J., Perera, R., Lanman, J., and Kuhn, R. J. (2014) Ultrastructural characterization and three-dimensional architecture of replication sites in Dengue virus-infected mosquito cells. *J. Virol.* **88**, 4687-4697
27. Richter, M. K. S., da Silva Voorham, J. M., Torres Pedraza, S., Hoornweg, T. E., van de Pol, D. P. I., Rodenhuis-Zybert, I. A., Wilschut, J., and Smit, J. M. (2014) Immature Dengue virus is infectious in human immature dendritic cells via interaction with the receptor molecule DC-SIGN. *PLoS ONE* **9**, e98785

28. Kostyuchenko, V. A., Zhang, Q., Tan, J. L., Ng, T.-S., and Lok, S.-M. (2013) Immature and mature Dengue serotype 1 virus structures provide insight into the maturation process. *J. Virol.* **87**, 7700-7707
29. Stadler, K., Allison, S. L., Schlich, J., and Heinz, F. X. (1997) Proteolytic activation of tick-borne encephalitis virus by furin. *J. Virol.* **71**, 8475-8481
30. 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., Baker, T. S., and Strauss, J. H. (2002) Structure of dengue virus: implications for flavivirus organization, maturation, and fusion. *Cell* **108**, 717-725
31. Zhang, Y., Corver, J., Chipman, P. R., Zhang, W., Pletnev, S. V., Sedlak, D., Baker, T. S., Strauss, J. H., Kuhn, R. J., and Rossmann, M. G. (2003) Structures of immature flavivirus particles. *EMBO J.* **22**, 2604-2613
32. 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. Mol. Biol.* **10**, 907-912
33. Jones, C. T., Ma, L., Burgner, J. W., Groesch, T. D., Post, C. B., and Kuhn, R. J. (2003) Flavivirus capsid is a dimeric alpha-helical protein. *J. Virol.* **77**, 7143-7149
34. Ma, L., Jones, C. T., Groesch, T. D., Kuhn, R. J., and Post, C. B. (2004) Solution structure of Dengue virus capsid protein reveals another fold. *Proc. Natl. Acad. Sci. USA* **101**, 3414-3419
35. Wengler, G., Wengler, G., and Gross, H. J. (1978) Studies on virus-specific nucleic acids synthesized in vertebrate and mosquito cells infected with flaviviruses. *Virology* **89**, 423-437
36. 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
37. Villordo, S. M., and Gamarnik, A. V. (2009) Genome cyclization as strategy for flavivirus RNA replication. *Virus Res.* **139**, 230-239
38. Li, K., Phoo, W. W., and Luo, D. (2014) Functional interplay among the flavivirus NS3 protease, helicase, and cofactors. *Virol. Sin.* **29**, 74-85
39. Edgil, D., Polacek, C., and Harris, E. (2006) Dengue virus utilizes a novel strategy for translation initiation when cap-dependent translation is inhibited. *J. Virol.* **80**, 2976-2986
40. Komar, A. A., and Hatzoglou, M. (2011) Cellular IRES-mediated translation: The war of ITAFs in pathophysiological states. *Cell Cycle* **10**, 229-240

41. 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
42. Welsch, S., Miller, S., Romero-Brey, I., Merz, A., Bleck, C. K. E., Walther, P., Fuller, S. D., Antony, C., Krijnse-Locker, J., and Bartenschlager, R. (2009) Composition and three-dimensional architecture of the Dengue virus replication and assembly sites. *Cell Host & Microbe* **5**, 365-375
43. Miller, S., and Krijnse-Locker, J. (2008) Modification of intracellular membrane structures for virus replication. *Nat Rev Micro* **6**, 363-374
44. Uchil, P. D., and Satchidanandam, V. (2003) Architecture of the flaviviral replication complex: protease, nuclease, and detergents reveal encasement within double-layered membrane compartments. *J. Biol. Chem.* **278**, 24388-24398
45. Mackenzie, J. (2005) Wrapping things up about virus RNA replication. *Traffic* **6**, 967-977
46. Netherton, C., Moffat, K., Brooks, E., and Wileman, T. (2007) A guide to viral inclusions, membrane rearrangements, factories, and viroplasm produced during virus replication. in *Adv. Virus Res.*, Academic Press. pp 101-182
47. 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.* **583**, 691-696
48. Wengler, G., and Wengler, G. (1993) The NS3 nonstructural protein of flaviviruses contains an RNA triphosphatase activity. *Virology* **197**, 265-273
49. 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
50. Davidson, A. D. (2009) Chapter 2 New insights into flavivirus nonstructural protein 5. in *Adv. Virus Res.* (Karl Maramorosch, A. J. S., and Frederick, A. M. eds.), Academic Press. pp 41-101
51. 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
52. Issur, M., Geiss, B. J., Bougie, I., Picard-Jean, F., Despins, S., Mayette, J., Hobdey, S. E., and Bisaillon, M. (2009) The flavivirus NS5 protein is a true RNA guanylyltransferase that catalyzes a two-step reaction to form the RNA cap structure. *RNA* **15**, 2340-2350

53. Miller, S., Kastner, S., Krijnse-Locker, J., Bühler, 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
54. Yu, L., Takeda, K., and Markoff, L. (2013) Protein–protein interactions among West Nile non-structural proteins and transmembrane complex formation in mammalian cells. *Virology* **446**, 365-377
55. 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
56. Mackenzie, J. M., Jones, M. K., and Young, P. R. (1996) Immunolocalization of the Dengue virus nonstructural glycoprotein NS1 suggests a role in viral RNA replication. *Virology* **220**, 232-240
57. Muller, D. A., and Young, P. R. (2013) The flavivirus NS1 protein: Molecular and structural biology, immunology, role in pathogenesis and application as a diagnostic biomarker. *Antiviral Res.* **98**, 192-208
58. Cleaves, G. R., Ryan, T. E., and Walter Schlesinger, R. (1981) Identification and characterization of type 2 Dengue virus replicative intermediate and replicative form RNAs. *Virology* **111**, 73-83
59. Raviprakash, K., Sinha, M., Hayes, C. G., and Porter, K. R. (1998) Conversion of Dengue virus replicative form RNA (RF) to replicative intermediate (RI) by nonstructural proteins NS-5 and NS-3. *Am. J. Trop. Med. Hyg.* **58**, 90-95
60. Klema, V., Padmanabhan, R., and Choi, K. (2015) Flaviviral replication complex: Coordination between RNA synthesis and 5'-RNA capping. *Viruses* **7**, 2837
61. Chu, P. W. G., and Westaway, E. G. (1985) Replication strategy of Kunjin Virus: Evidence for recycling role of replicative form RNA as template in semiconservative and asymmetric replication. *Virology* **140**, 68-79
62. Ferron, F., Decroly, E., Selisko, B., and Canard, B. (2012) The viral RNA capping machinery as a target for antiviral drugs. *Antiviral Res.* **96**, 21-31
63. Alvarez, D. E., De Lella Ezcurra, A. L., Fucito, S., and Gamarnik, A. V. (2005) Role of RNA structures present at the 3' UTR of dengue virus on translation, RNA synthesis, and viral replication. *Virology* **339**, 200-212
64. Filomatori, C. V., Iglesias, N. G., Villordo, S. M., Alvarez, D. E., and Gamarnik, A. V. (2011) RNA sequences and structures required for the recruitment and activity of the Dengue virus polymerase. *J. Biol. Chem.* **286**, 6929-6939
65. Cui, T., Sugrue, R. J., Xu, Q., Lee, A. K. W., 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
66. Alvarez, D. E., Lodeiro, M. F., Ludueña, S. J., Pietrasanta, L. I., and Gamarnik, A. V. (2005) Long-range RNA-RNA interactions circularize the Dengue virus genome. *J. Virol.* **79**, 6631-6643
67. 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
68. Potisopon, S., Priet, S., Collet, A., Decroly, E., Canard, B., and Selisko, B. (2014) The methyltransferase domain of Dengue virus protein NS5 ensures efficient RNA synthesis initiation and elongation by the polymerase domain. *Nucleic Acids Res.* **42**, 11642-11656
69. Kapoor, M., Zhang, L., Ramachandra, M., Kusakawa, J., Ebner, K. E., and Padmanabhan, R. (1995) Association between NS3 and NS5 proteins of Dengue virus type 2 in the putative RNA replicase II linked to differential phosphorylation of NS5. *J. Biol. Chem.* **270**, 19100-19106
70. 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
71. 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 β 1 and importin α/β -recognized nuclear localization signals. *J. Biol. Chem.* **277**, 36399-36407
72. Kumar, A., Bühler, S., Selisko, B., Davidson, A., Mulder, K., Canard, B., Miller, S., and Bartenschlager, R. (2013) Nuclear localization of Dengue virus nonstructural protein 5 does not strictly correlate with efficient viral RNA replication and inhibition of type I interferon signaling. *J. Virol.* **87**, 4545-4557
73. Hannemann, H., Sung, P.-Y., Chiu, H.-C., Yousuf, A., Bird, J., Lim, S. P., and Davidson, A. D. (2013) Serotype-specific differences in Dengue virus non-structural protein 5 nuclear localization. *J. Biol. Chem.* **288**, 22621-22635
74. Tay, M. Y. F., Fraser, J. E., Chan, W. K. K., Moreland, N. J., Rathore, A. P., Wang, C., Vasudevan, S. G., and Jans, D. A. (2013) Nuclear localization of dengue virus (DENV) 1–4 non-structural protein 5; protection against all 4 DENV serotypes by the inhibitor Ivermectin. *Antiviral Res.* **99**, 301-306
75. Mazzon, M., Jones, M., Davidson, A., Chain, B., and Jacobs, M. (2009) Dengue virus NS5 inhibits interferon- α signaling by blocking signal transducer and activator of transcription 2 phosphorylation. *J. Infect. Dis.* **200**, 1261-1270

-
76. Ashour, J., Laurent-Rolle, M., Shi, P.-Y., and García-Sastre, A. (2009) NS5 of Dengue virus mediates STAT2 binding and degradation. *J. Virol.* **83**, 5408-5418
77. Ashour, J., Morrison, J., Laurent-Rolle, M., Belicha-Villanueva, A., Plumlee, C. R., Bernal-Rubio, D., Williams, K. L., Harris, E., Fernandez-Sesma, A., Schindler, C., and García-Sastre, A. (2010) Mouse STAT2 restricts early Dengue virus replication. *Cell Host & Microbe* **8**, 410-421
78. Cheng, X., and Roberts, R. J. (2001) AdoMet-dependent methylation, DNA methyltransferases and base flipping. *Nucleic Acids Res.* **29**, 3784-3795
79. Lim, S. P., Sonntag, L. S., Noble, C., Nilar, S. H., Ng, R. H., Zou, G., Monaghan, P., Chung, K. Y., Dong, H., Liu, B., Bodenreider, C., Lee, G., Ding, M., Chan, W. L., Wang, G., Jian, Y. L., Chao, A. T., Lescar, J., Yin, Z., Vedananda, T. R., Keller, T. H., and Shi, P.-Y. (2011) Small molecule inhibitors that selectively block Dengue virus methyltransferase. *J. Biol. Chem.* **286**, 6233-6240
80. Dong, H., Chang, D. C., Xie, X., Toh, Y. X., Chung, K. Y., Zou, G., Lescar, J., Lim, S. P., and Shi, P.-Y. (2010) Biochemical and genetic characterization of Dengue virus methyltransferase. *Virology* **405**, 568-578
81. 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
82. Bhattacharya, D., Hoover, S., Falk, S. P., Weisblum, B., Vestling, M., and Striker, R. (2008) Phosphorylation of yellow fever virus NS5 alters methyltransferase activity. *Virology* **380**, 276-284
83. Zhao, Y., Soh, T. S., Lim, S. P., Chung, K. Y., Swaminathan, K., Vasudevan, S. G., Shi, P.-Y., Lescar, J., and Luo, D. (2015) Molecular basis for specific viral RNA recognition and 2' -O-ribose methylation by the Dengue virus nonstructural protein 5 (NS5). *Proc. Natl. Acad. Sci.* **112**, 14834-14839
84. Zhao, Y., Soh, T. S., Zheng, J., Chan, K. W. K., Phoo, W. W., Lee, C. C., Tay, M. Y. F., Swaminathan, K., Cornvik, T. C., Lim, S. P., Shi, P.-Y., Lescar, J., Vasudevan, S. G., and Luo, D. (2015) A crystal structure of the Dengue virus NS5 protein reveals a novel inter-domain interface essential for protein flexibility and virus replication. *PLoS Pathog* **11**, e1004682
85. Drake, J. W. (1999) The distribution of rates of spontaneous mutation over viruses, prokaryotes, and eukaryotes. *Ann. N. Y. Acad. Sci.* **870**, 100-107
86. Lim, S. P., Koh, J. H. K., Seh, C. C., Liew, C. W., Davidson, A. D., Chua, L. S., Chandrasekaran, R., Cornvik, T. C., Shi, P.-Y., and Lescar, J. (2013) A crystal structure of the Dengue virus non-structural protein 5 (NS5) polymerase delineates interdomain amino acid residues that enhance its thermostability and *de Novo* initiation activities. *J. Biol. Chem.* **288**, 31105-31114

87. Boehr, D., Arnold, J., Moustafa, I., and Cameron, C. (2014) Structure, dynamics, and fidelity of RNA-dependent RNA polymerases. in *Nucleic Acid Polymerases* (Murakami, K. S., and Trakselis, M. A. eds.), Springer Berlin Heidelberg. pp 309-333
88. Choi, K. (2012) Viral Polymerases. in *Viral Molecular Machines* (Rossmann, M. G., and Rao, V. B. eds.), Springer US. pp 267-304
89. Poch, O., Sauvaget, I., Delarue, M., and Tordo, N. (1989) Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J.* **8**, 3867-3874
90. 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- β and the viral helicase, NS3. *J. Gen. Virol.* **82**, 735-745
91. Tay, M. Y. F., Saw, W. G., Zhao, Y., Chan, K. W. K., Singh, D., Chong, Y., Forwood, J. K., Ooi, E. E., Grüber, G., Lescar, J., Luo, D., and Vasudevan, S. G. (2015) The C-terminal 50 amino acid residues of Dengue NS3 protein are important for NS3-NS5 interaction and viral replication. *J. Biol. Chem.* **290**, 2379-2394
92. Teramoto, T., Boonyasuppayakorn, S., Handley, M., Choi, K. H., and Padmanabhan, R. (2014) Substitution of NS5 N-terminal domain of Dengue virus type 2 RNA with type 4 domain caused impaired replication and emergence of adaptive mutants with enhanced fitness. *J. Biol. Chem.* **289**, 22385-22400
93. Selisko, B., Dutartre, H., Guillemot, J.-C., Debarnot, C., Benarroch, D., Khromykh, A., Desprès, P., Egloff, M.-P., and Canard, B. (2006) Comparative mechanistic studies of de novo RNA synthesis by flavivirus RNA-dependent RNA polymerases. *Virology* **351**, 145-158
94. Lu, G., and Gong, P. (2013) Crystal structure of the full-length Japanese Encephalitis Virus NS5 reveals a conserved methyltransferase-polymerase interface. *PLoS Pathog* **9**, e1003549
95. Zhao, Y., Soh, T. S., Chan, K. W. K., Fung, S. S. Y., Swaminathan, K., Lim, S. P., Shi, P.-Y., Huber, T., Lescar, J., Luo, D., and Vasudevan, S. G. (2015) Flexibility of NS5 methyltransferase-polymerase linker region is essential for Dengue virus replication. *J. Virol.* **89**, 10717-10721
96. Zou, G., Chen, Y.-L., Dong, H., Lim, C. C., Yap, L. J., Yau, Y. H., Shochat, S. G., Lescar, J., and Shi, P.-Y. (2011) Functional analysis of two cavities in flavivirus NS5 polymerase. *J. Biol. Chem.* **286**, 14362-14372
97. Kroschewski, H., Lim, S. P., Butcher, R. E., Yap, T. L., Lescar, J., Wright, P. J., Vasudevan, S. G., and Davidson, A. D. (2008) Mutagenesis of the Dengue virus type 2 NS5 methyltransferase domain. *J. Biol. Chem.* **283**, 19410-19421

-
98. Takahashi, H., Takahashi, C., Moreland, N. J., Chang, Y.-T., Sawasaki, T., Ryo, A., Vasudevan, S. G., Suzuki, Y., and Yamamoto, N. (2012) Establishment of a robust Dengue virus NS3–NS5 binding assay for identification of protein–protein interaction inhibitors. *Antiviral Res.* **96**, 305–314
 99. Luo, D., Wei, N., Doan, D. N., Paradkar, P. N., Chong, Y., Davidson, A. D., Kotaka, M., Lescar, J., and Vasudevan, S. G. (2010) Flexibility between the protease and helicase domains of the Dengue virus NS3 protein conferred by the linker region and its functional implications. *J. Biol. Chem.* **285**, 18817–18827
 100. Luo, D., Xu, T., Hunke, C., Grüber, G., Vasudevan, S. G., and Lescar, J. (2008) Crystal structure of the NS3 protease-helicase from Dengue virus. *J. Virol.* **82**, 173–183
 101. 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
 102. Ramanathan, M. P., Chambers, J. A., Pankhong, P., Chattergoon, M., Attatippaholkun, W., Dang, K., Shah, N., and Weiner, D. B. (2006) Host cell killing by the West Nile Virus NS2B–NS3 proteolytic complex: NS3 alone is sufficient to recruit caspase-8-based apoptotic pathway. *Virology* **345**, 56–72
 103. Shafee, N., and AbuBakar, S. (2003) Dengue virus type 2 NS3 protease and NS2B–NS3 protease precursor induce apoptosis. *J. Gen. Virol.* **84**, 2191–2195
 104. Heaton, N. S., Perera, R., Berger, K. L., Khadka, S., LaCount, D. J., Kuhn, R. J., and Randall, G. (2010) Dengue virus nonstructural protein 3 redistributes fatty acid synthase to sites of viral replication and increases cellular fatty acid synthesis. *Proc. Natl. Acad. Sci.* **107**, 17345–17350
 105. Tang, W.-C., Lin, R.-J., Liao, C.-L., and Lin, Y.-L. (2014) Rab18 facilitates Dengue virus infection by targeting fatty acid synthase to sites of viral replication. *J. Virol.* **88**, 6793–6804
 106. Chambers, T. J., Hahn, C. S., Galler, R., and Rice, C. M. (1990) Flavivirus genome organization, expression, and replication. *Annu. Rev. Microbiol.* **44**, 649–688
 107. 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
 108. 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

109. Noble, C. G., Seh, C. C., Chao, A. T., and Shi, P. Y. (2012) Ligand-bound structures of the Dengue virus protease reveal the active conformation. *J. Virol.* **86**, 438-446
110. 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
111. Kim, Y. M., Gayen, S., Kang, C., Joy, J., Huang, Q., Chen, A. S., Wee, J. L. K., Ang, M. J. Y., Lim, H. A., Hung, A. W., Li, R., Noble, C. G., Lee, L. T., Yip, A., Wang, Q.-Y., Chia, C. S. B., Hill, J., Shi, P.-Y., and Keller, T. H. (2013) NMR analysis of a novel enzymatically active unlinked Dengue NS2B-NS3 protease complex. *J. Biol. Chem.* **288**, 12891-12900
112. Phong, Wai Y., Moreland, Nicole J., Lim, Siew P., Wen, D., Paradkar, Prasad N., and Vasudevan, Subhash G. (2011) Dengue protease activity: the structural integrity and interaction of NS2B with NS3 protease and its potential as a drug target. *Biosci. Rep.* **31**, 399-409
113. 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
114. 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
115. 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., Vasudevan, S. G., and Lescar, J. (2008) Insights into RNA unwinding and ATP hydrolysis by the flavivirus NS3 protein. *EMBO J.* **27**, 3209-3219
116. Luo, D., Vasudevan, S. G., and Lescar, J. (2015) The flavivirus NS2B-NS3 protease-helicase as a target for antiviral drug development. *Antiviral Res.* **118**, 148-158
117. Coutard, B., Decroly, E., Li, C., Sharff, A., Lescar, J., Bricogne, G., and Barral, K. (2014) Assessment of Dengue virus helicase and methyltransferase as targets for fragment-based drug discovery. *Antiviral Res.* **106**, 61-70
118. Castro, C., Arnold, J. J., and Cameron, C. E. (2005) Incorporation fidelity of the viral RNA-dependent RNA polymerase: a kinetic, thermodynamic and structural perspective. *Virus Res.* **107**, 141-149

119. Crotty, S., Cameron, C. E., and Andino, R. (2001) RNA virus error catastrophe: Direct molecular test by using ribavirin. *Proc. Natl. Acad. Sci.* **98**, 6895-6900
120. Deore, R. R., and Chern, J. W. (2010) NS5B RNA dependent RNA polymerase inhibitors: The promising approach to treat Hepatitis C virus infections. *Curr. Med. Chem.* **17**, 3806-3826
121. Lim, S. P., Wang, Q.-Y., Noble, C. G., Chen, Y.-L., Dong, H., Zou, B., Yokokawa, F., Nilar, S., Smith, P., Beer, D., Lescar, J., and Shi, P.-Y. (2013) Ten years of dengue drug discovery: Progress and prospects. *Antiviral Res.* **100**, 500-519
122. Mastrangelo, E., Pezzullo, M., De Burghgraeve, T., Kaptein, S., Pastorino, B., Dallmeier, K., de Lamballerie, X., Neyts, J., Hanson, A. M., Frick, D. N., Bolognesi, M., and Milani, M. (2012) Ivermectin is a potent inhibitor of flavivirus replication specifically targeting NS3 helicase activity: new prospects for an old drug. *J. Antimicrob. Chemother.* **67**, 1884-1894
123. López, C., Gil, L., Lazo, L., Menéndez, I., Marcos, E., Sánchez, J., Valdés, I., Falcón, V., de la Rosa, M., Márquez, G., Guillén, G., and Hermida, L. (2009) *In vitro* assembly of nucleocapsid-like particles from purified recombinant capsid protein of Dengue-2 virus. *Arch. Virol.* **154**, 695-698
124. Samsa, M. M., Mondotte, J. A., Iglesias, N. G., Assunção-Miranda, I., Barbosa-Lima, G., Da Poian, A. T., Bozza, P. T., and Gamarnik, A. V. (2009) Dengue virus capsid protein usurps lipid droplets for viral particle formation. *PLoS Pathog* **5**, e1000632
125. Faustino, A. F., Carvalho, F. A., Martins, I. C., Castanho, M. A. R. B., Mohana-Borges, R., Almeida, F. C. L., Da Poian, A. T., and Santos, N. C. (2014) Dengue virus capsid protein interacts specifically with very low-density lipoproteins. *Nanomed. Nanotechnol. Biol. Med.* **10**, 247-255
126. Faustino, A. F., Martins, I. C., Carvalho, F. A., Castanho, M. A. R. B., Maurer-Stroh, S., and Santos, N. C. (2015) Understanding Dengue virus capsid protein interaction with key biological targets. *Sci. Rep.* **5**, 10592
127. Faustino, A. F., Guerra, G. M., Huber, R. G., Hollmann, A., Domingues, M. M., Barbosa, G. M., Enguita, F. J., Bond, P. J., Castanho, M. A. R. B., Da Poian, A. T., Almeida, F. C. L., Santos, N. C., and Martins, I. C. (2015) Understanding Dengue virus capsid protein disordered N-terminus and pep14-23-based inhibition. *ACS Chem. Biol.* **10**, 517-526
128. Carvalho, F. A., Carneiro, F. A., Martins, I. C., Assunção-Miranda, I., Faustino, A. F., Pereira, R. M., Bozza, P. T., Castanho, M. A. R. B., Mohana-Borges, R., Da Poian, A. T., and Santos, N. C. (2012) Dengue virus capsid protein binding to hepatic lipid droplets (LD) is potassium ion dependent and is mediated by LD surface proteins. *J. Virol.* **86**, 2096-2108

129. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res.* **28**, 235-242
130. Putnam, C. D., Hammel, M., Hura, G. L., and Tainer, J. A. (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys* **40**, 191-285
131. Schmidt, A., and Lamzin, V. S. (2007) From atoms to proteins. *Cell. Mol. Life Sci.* **64**, 1959-1969
132. Bernadó, P., Mylonas, E., Petoukhov, M. V., Blackledge, M., and Svergun, D. I. (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* **129**, 5656-5664
133. Tria, G., Mertens, H. D., Kachala, M., and Svergun, D. I. (2015) Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ* **2**, 207-217
134. Bu, Z., and Callaway, D. J. E. (2011) Chapter 5 - Proteins MOVE! Protein dynamics and long-range allostery in cell signaling. in *Advances in Protein Chemistry and Structural Biology* (Rossen, D. ed.), Academic Press. pp 163-221
135. Shamir, M., Bar-On, Y., Phillips, R., and Milo, R. (2016) SnapShot: Timescales in Cell Biology. *Cell* **164**, 1302-1302.e1301
136. Koch, M. H., Vachette, P., and Svergun, D. I. (2003) Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Q Rev Biophys* **36**, 147-227
137. Matsui, T., Tsuruta, H., and Johnson, J. E. (2010) Balanced electrostatic and structural forces guide the large conformational change associated with maturation of T = 4 virus. *Biophys. J.* **98**, 1337-1343
138. Graceffa, R., Nobrega, R. P., Barrea, R. A., Kathuria, S. V., Chakravarthy, S., Bilsel, O., and Irving, T. C. (2013) Sub-millisecond time-resolved SAXS using a continuous-flow mixer and X-ray microbeam. *J. Synchrotron Radiat.* **20**, 820-825
139. Doniach, S. (2001) Changes in biomolecular conformation seen by small angle X-ray scattering. *Chemical Reviews* **101**, 1763-1778
140. Ortore, M. G., Spinozzi, F., Vilasi, S., Sirangelo, I., Irace, G., Shukla, A., Narayanan, T., Sinibaldi, R., and Mariani, P. (2011) Time-resolved small-angle x-ray scattering study of the early stage of amyloid formation of an apomyoglobin mutant. *Phys. Rev. E* **84**, 061904
141. Russell, R., Millett, I. S., Tate, M. W., Kwok, L. W., Nakatani, B., Gruner, S. M., Mochrie, S. G. J., Pande, V., Doniach, S., Herschlag, D., and Pollack, L.

- (2002) Rapid compaction during RNA folding. *Proc. Natl. Acad. Sci.* **99**, 4266-4271
142. Kler, S., Asor, R., Li, C., Ginsburg, A., Harries, D., Oppenheim, A., Zlotnick, A., and Raviv, U. (2012) RNA Encapsidation by SV40-Derived Nanoparticles Follows a Rapid Two-State Mechanism. *J. Am. Chem. Soc.* **134**, 8823-8830
143. Tresset, G., Le Coeur, C., Bryche, J.-F., Tatou, M., Zeghal, M., Charpilienne, A., Poncet, D., Constantin, D., and Bressanelli, S. (2013) Norovirus capsid proteins self-assemble through biphasic kinetics via long-lived state-like intermediates. *J. Am. Chem. Soc.* **135**, 15373-15381
144. Cao, E., Chen, Y., Cui, Z., and Foster, P. R. (2003) Effect of freezing and thawing rates on denaturation of proteins in aqueous solutions. *Biotechnol. Bioeng.* **82**, 684-690
145. Franks, F. (1985) *Biophysics and biochemistry at low temperatures*, Cambridge University Press, Cambridge Cambridgeshire ; New York
146. Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685
147. Smolsky, I. L., Liu, P., Niebuhr, M., Ito, K., Weiss, T. M., and Tsuruta, H. (2007) Biological small-angle X-ray scattering facility at the Stanford Synchrotron Radiation Laboratory. *J. Appl. Crystallogr.* **40**, s453-s458
148. Martel, A., Liu, P., Weiss, T. M., Niebuhr, M., and Tsuruta, H. (2012) An integrated high-throughput data acquisition system for biological solution X-ray scattering studies. *J. Synchrotron Radiat.* **19**, 431-434
149. Tsuruta, H., Brennan, S., Rek, Z. U., Irving, T. C., Tompkins, W. H., and Hodgson, K. O. (1998) A wide-bandpass multilayer monochromator for biological small-angle scattering and fiber diffraction studies. *J. Appl. Crystallogr.* **31**, 672-682
150. Konarev, P. V., Petoukhov, M. V., Volkov, V. V., and Svergun, D. I. (2006) ATSAS 2.1, a program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **39**, 277-286
151. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277-1282
152. Jacques, D. A., and Trewhella, J. (2010) Small-angle scattering for structural biology—Expanding the frontier while avoiding the pitfalls. *Protein Science : A Publication of the Protein Society* **19**, 642-657
153. Guinier, A. (1939) Diffraction of X-rays of very small angles-application to the study of ultramicroscopic phenomenon. *Ann. Phys. (Paris)* **12**, 161-237
154. Gunier, A., and Fournet, G. (1955) *Small angle scattering of X-rays*, Wiley, New York

155. Petoukhov, M. V., Konarev, P. V., Kikhney, A. G., and Svergun, D. I. (2007) ATSAS 2.1 - towards automated and web-supported small-angle scattering data analysis. *J. Appl. Crystallogr.* **40**, s223-s228
156. Svergun, D. I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495-503
157. Glatter, O. (1977) Data evaluation in small angle scattering: calculation of the radial electron density distribution by means of indirect Fourier transformation. *Acta Physica Austriaca* **47**, 83-102
158. Franke, D., and Svergun, D. I. (2009) DAMMIF, a program for rapid *ab-initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42**, 342-346
159. Svergun, D. I., Petoukhov, M. V., and Koch, M. H. J. (2001) Determination of domain structure of proteins from X-Ray solution scattering. *Biophys. J.* **80**, 2946-2953
160. Volkov, V. V., and Svergun, D. I. (2003) Uniqueness of *ab initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36**, 860-864
161. Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D. T., Konarev, P. V., and Svergun, D. I. (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45**, 342-350
162. Durand, D., Vivès, C., Cannella, D., Pérez, J., Pebay-Peyroula, E., Vachette, P., and Fieschi, F. (2010) NADPH oxidase activator p67phox behaves in solution as a multidomain protein with semi-flexible linkers. *J. Struct. Biol.* **169**, 45-53
163. Receveur-Brechot, V., and Durand, D. (2012) How random are intrinsically disordered proteins? A small angle scattering perspective. *Curr Protein Pept Sci* **13**, 55-75
164. Reyes, F. E., Schwartz, C. R., Tainer, J. A., and Rambo, R. P. (2014) Chapter 11 - Methods for using new conceptual tools and parameters to assess RNA structure by small-angle X-Ray scattering. in *Methods Enzymol.* (Donald, H. B.-A. ed.), Academic Press. pp 235-263
165. Rambo, R. P., and Tainer, J. A. (2011) Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers* **95**, 559-571
166. Petoukhov, M. V., and Svergun, D. I. (2005) Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J.* **89**, 1237-1250
167. Svergun, D., Barberato, C., and Koch, M. H. J. (1995) CRY SOL - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768-773

-
168. Kozin, M. B., and Svergun, D. I. (2001) Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr.* **34**, 33-41
169. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191
170. Pickett, B. E., Sadat, E. L., Zhang, Y., Noronha, J. M., Squires, R. B., Hunt, V., Liu, M., Kumar, S., Zaremba, S., Gu, Z., Zhou, L., Larson, C. N., Dietrich, J., Klem, E. B., and Scheuermann, R. H. (2012) ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic Acids Res.* **40**, D593-D598
171. Consortium, T. U. (2012) Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* **40**, D71-D75
172. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539
173. Saw, W. G., Tria, G., Grüber, A., Subramanian Manimekalai, M. S., Zhao, Y., Chandramohan, A., Srinivasan Anand, G., Matsui, T., Weiss, T. M., Vasudevan, S. G., and Grüber, G. (2015) Structural insight and flexible features of NS5 proteins from all four serotypes of Dengue virus in solution. *Acta Crystallogr D* **71**, 2309-2327
174. Poch, O., Sauvaget, I., Delarue, M., and Tordo, N. (1989) Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J.* **8**, 3867-3874
175. Bernado, P., and Svergun, D. I. (2012) Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Molecular BioSystems* **8**, 151-167
176. Subramanian Manimekalai, M. S., Saw, W. G., Pan, A., Grüber, A., and Grüber, G. (2016) Identification of the critical linker residues conferring differences in the compactness of NS5 from Dengue virus serotype 4 and NS5 from Dengue virus serotypes 1-3. *Acta Crystallogr D* **72**, 795-807
177. Henderson, B. R., Saeedi, B. J., Campagnola, G., and Geiss, B. J. (2011) Analysis of RNA binding by the Dengue virus NS5 RNA capping enzyme. *PLoS ONE* **6**, e25795
178. Assenberg, R., Mastrangelo, E., Walter, T. S., Verma, A., Milani, M., Owens, R. J., Stuart, D. I., Grimes, J. M., and Mancini, E. J. (2009) Crystal structure of a novel conformational state of the flavivirus NS3 protein: Implications for polyprotein processing and viral replication. *J. Virol.* **83**, 12895-12906
179. 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
180. Bussetta, C., and Choi, K. H. (2012) Dengue virus nonstructural protein 5 adopts multiple conformations in solution. *Biochemistry* **51**, 5921-5931
181. Malet, H., Egloff, M.-P., Selisko, B., Butcher, R. E., Wright, P. J., Roberts, M., Gruez, A., Sulzenbacher, G., Vonnrhein, C., Bricogne, G., Mackenzie, J. M., Khromykh, A. A., Davidson, A. D., and Canard, B. (2007) Crystal structure of the RNA polymerase domain of the West Nile virus non-structural protein 5. *J. Biol. Chem.* **282**, 10678-10689
182. 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
183. 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
184. Markoff, L., Falgout, B., and Chang, A. (1997) A conserved internal hydrophobic domain mediates the stable membrane integration of the Dengue virus capsid protein. *Virology* **233**, 105-117

Author's publication related to the projects

1. Tay, M. Y. F., **Saw, W. G.**, Zhao, Y., Chan, W. K. K., Singh, D, Chong, Y., Forwood, J. K., Ooi, E. E., Grüber, G., Lescar, J., Luo, D., and Vasudevan, S. G. (2015) The C-terminal 50 amino acid residues of Dengue NS3 protein are important for NS3-NS5 interaction and viral replication. *J. Biol. Chem.* **290**, 2379-2394
2. **Saw, W. G.**, Tria, G., Grüber, A., Manimekalai, M. S. S., Zhao, Y., Chandramohan, A., Anand, G. S., Matsui, T., Weiss, T., Vasudevan, S. and Grüber, G. (2015) Structural insight and flexibility features of NS5 proteins from all four serotypes of Dengue virus in solution. *Acta Crystallogr. D* **71**, 2309-2327
3. Manimekalai, M.S.S., **Saw, W. G.**, Pan, A., Grüber, A. and Grüber, G. (2016) Identification of the critical linker residues conferring differences in compactness of DENV-4 NS5 from Dengue virus serotypes 1-3. *Acta Crystallogr. D* **72**, 795-807

Conference attendance

1. 7th Asia-Pacific Organization for Cell Biology Congress and American Society for Cell Biology (APOCB), Singapore from 24th – 27th Feb 2014
2. Multiscale Dynamics Dengue Virus Conference, Singapore from 12th – 13th November 2014
3. 16th International Conference on Small-Angle Scattering (SAS-2015), Berlin, Germany from 13th – 18th September 2015

Posters and abstracts

1. **Saw, W. G.**, Eisenhaber, F. and Grüber, G. (2014) “*Structural studies on yGPAAL from Saccharomyces cerevisiae GPI Transamidase.*” 7th Asia-Pacific Organization for Cell Biology Congress and American Society for Cell Biology
2. **Saw, W. G.**, Tria, G., Grüber, A., Vasudevan, S. and Grüber, G. (2015) “*Dynamic features of the Dengue virus NS5 protein of serotype 3 in solution.*” in 16th International Conference on Small-Angle Scattering (SAS-2015), Berlin, Germany

Awards

Nanyang research scholarship (2012-2016)