STRUCTURE OF THE DENGUE VIRUS
HELICASE/NUCLEOSIDE TRIPHOSPHATASE
CATALYTIC DOMAIN

Xu Ting

SCHOOL OF BIOLOGICAL SCIENCES
NANYANG TECHNOLOGICAL UNIVERSITY

2007
ACKNOWLEDGEMENTS

First and foremost, I am deeply grateful to my supervisor, Dr. Julien Lescar, for giving me this great opportunity to learn crystallography; encouragement and support he has given me throughout my research and the preparation of this thesis.

Next, I wish to express my sincere gratitude to Dr. Subhash G.Vasudevan, head of dengue unit of Novartis Institute of Tropical Diseases (NITD), for his initiation of the project of crystal structure determination of dengue NS3 helicase domain.

I would also like to thank Daying Wen, Alex Chao, (NITD) for their sincere help in this research and special thanks to Dr. Aruna Sampath (NITD) for sharing the biochemical data which made our publication more powerful. I owe my sincere thanks to Dr. Max Nanao (European Synchrotron Radiation Facility) for his great help in data collection and valuable suggestions in structure determination.

Thanks also go to my labmates, Dr. Amy Ooi, Ravikumar Rajamanonmani, Mo Min, Fan Hui, Yap Thai Leong & Lee Hooi Chen, for their support and creating a friendly atmosphere which made the days in the lab more bearable and enjoyable.

Finally, I wish to extend my gratitude to my parents and my wife. This thesis could not have been done without their constant encouragement and endless love.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS.................................................................1

TABLE OF CONTENTS.............................................................2

SUMMARY.................................................................................6

PUBLICATIONS........................................................................8

LIST OF TABLES......................................................................9

LIST OF FIGURES...................................................................10

ABBREVIATIONS....................................................................13

CHAPTER 1 INTRODUCTION..................................................17

1.1 Viruses..............................................................................16

1.2 Classification of viruses....................................................17

1.3 *Flaviviridae*.....................................................................19

1.4 Flaviviruses.......................................................................19

1.5 Dengue and dengue virus classification..............................20

1.6 Pathogenesis of dengue virus.............................................22

1.7 Replication cycle of flaviviruses.........................................24

1.8 Structure of dengue virus...................................................25

1.9 Dengue virus proteins.......................................................28

1.9.1 Structural proteins.......................................................29

1.9.1.1 Capsid protein (C)..................................................30

1.9.1.2 Membrane protein (prM and M)..............................32

1.9.1.3 Envelope protein (E)............................................34

1.9.2 Non-structural proteins.................................................36
# Table of contents

1.9.2.1 NS1 ................................................................................................. 36
1.9.2.2 NS2A/2B ......................................................................................... 38
1.9.2.3 NS3 ................................................................................................. 40
  1.9.2.3.1 NS3 protease domain (NS3pro) .................................................. 40
  1.9.2.3.2 NS3 helicase domain (NS3hel) .................................................. 43
    1.9.2.3.2.1 NTPase/helicase activity ......................................................... 44
    1.9.2.3.2.2 RTPase activity ......................................................................... 46
1.9.2.4 NS4A/4B ......................................................................................... 47
1.9.2.5 NS5 ................................................................................................. 47
1.10 Helicases: An ATP-fueled molecular motor ........................................... 51
  1.10.1 Helicases classification ................................................................... 52
    1.10.1.1 SF1 & SF2 superfamilies ............................................................. 55
      1.10.1.1.1 Helicase motifs ....................................................................... 56
    1.10.1.2 SF3 superfamily ........................................................................... 58
    1.10.1.3 Family 4 & 5 ................................................................................ 59
  1.10.2 Helicase structures and mechanisms ............................................... 59
    1.10.2.1 Structures and mechanisms of SF1 helicases ......................... 63
    1.10.2.2 Structures and putative mechanisms of SF2 helicases .......... 67
    1.10.2.3 Structures and mechanisms of SF3 helicases ....................... 73
    1.10.2.4 Structures and mechanisms of F4 helicases ......................... 77
1.11 Aims of the project............................................................................... 80

# CHAPTER 2 MATERIALS AND METHODS ............................................... 81

2.1 Overview ............................................................................................. 81
2.2 Cloning of DNS3H ............................................................................. 81
  2.2.1 Cloning strategy .............................................................................. 81
# Table of contents

2.2.2 PCR, restriction enzyme digestion and ligation...............................83  
2.2.3 Construct confirmation .................................................................84  
2.3 Expression of Trx-DNS3H ........................................................................84  
2.4 Purification of DNS3H .............................................................................85  
2.4.1 Overview..........................................................................................85  
2.4.2 Immobilized metal affinity chromatography (IMAC) .....................86  
2.4.3 Anion exchange chromatography ....................................................87  
2.4.4 Enterokinase cleavage of Trx-DNS3H ............................................88  
2.4.5 Removal of the thioredoxin protein tag ...........................................90  
2.4.6 Gel filtration chromatography..........................................................90  
2.5 Expression of Selenomethionine DNS3H.................................................91  
2.6 Purification of Selenomethionine DNS3H.............................................93  
2.7 Mass spectrometry of native and SeMet DNS3H .....................................93  
2.8 Phasing (Single-wavelength Anomalous Dispersion) ..............................94  
2.8.1 Phase ambiguity intrinsic to SAD....................................................94  
2.8.2 Bimodal phase probability distribution from SAD..........................97  
2.8.3 Phase improvement..........................................................................98  

CHAPTER 3 RESULTS AND DISCUSSION .................................................99  

3.1 Crystallization of DNS3H.........................................................................99  
3.2 Data collection and reduction .................................................................100  
3.3 Phasing of DNS3H..................................................................................102  
3.4 Model building, refinements and quality of the model...........................103  
3.5 The structure of DNS3H.........................................................................107  
3.5.1 Overall structure.............................................................................107  
3.5.2 Non-crystallographic dimer..............................................................113
### Table of contents

3.5.3 The NTPase active site ........................................................................ 116

3.5.4 Nucleic acid binding sites ............................................................... 121

3.6 Helicase mechanism ......................................................................... 128

3.7 Helicase mechanism of DNS3H ......................................................... 131

3.8 The full length NS3 protein ............................................................... 133

3.8 Replication complex ........................................................................ 135

3.9 Conclusion and perspectives ............................................................ 136

REFERENCES ....................................................................................... 138

APPENDIX ............................................................................................. 168
SUMMARY

Dengue fever is an important emerging public health concern, with several million viral infections occurring annually, for which no effective therapy currently exists. The NS3 protein of Dengue virus is a multifunctional protein endowed with protease, helicase, nucleoside 5’-triphosphatase (NTPase) and RNA triphosphatase activities. Therefore, NS3 plays an important role in viral life cycle and represents a very interesting target for the development of specific antiviral drug design. This thesis reports the structure of an enzymatically active fragment of the Dengue virus NTPase/helicase catalytic domain. The structure is composed of three domains, with two parallel “RecA-like” cores like in the HCV helicase and one additional domain composed mainly of α-helices which is strikingly different compared to that of hepatitis C virus NS3 helicase. A significant feature of this structure is a tunnel at its center surrounded by residues originating from the three domains, which is large enough to accommodate single-stranded RNA. The bound manganese and sulfate ions inside NTPase binding pocket reveals residues involved in the divalent metal-dependent NTPase catalytic mechanism. Comparison with the hepatitis C virus NS3 helicase complexed to single-stranded DNA (PDB ID: 1A1V, (Kim et al., 1998)) would place the 3’ single-stranded tail of a nucleic acid duplex in the tunnel that runs across the basic face of the protein. A sulfate ion is found to be situated inside the nucleic acid binding tunnel, which is close to the phosphodiester backbone of the single-stranded DNA in 1A1V. Comparison with the yellow fever virus NS3 helicase (Wu et al., 2005) reveals rigid body movements in domain 3 and a conformational change in motif V, which results in a surface charge switch inside the nucleic acid binding tunnel. The
Summary

difference between dengue and yellow fever NS3 helicases together with previous biochemical studies in hepatitis C virus NS3 helicase (Levin et al., 2003; 2005) provides evidence for the “Brownian motor” mechanism, in which the unidirectional translocation is fueled by single-stranded DNA binding and ATP binding allows for a brief period of random movement that prepares the helicase for the next cycle. A possible model accounting for dengue NS3 helicase activity is then proposed.
PUBLICATIONS

Publication related to this thesis:


Other publications:

LIST OF TABLES

Table 1. 1 Classification of viruses .................................................................18
Table 1. 2 Representatives of helicases ............................................................61
Table 2. 1 PCR reaction for amplification of DNS3H gene ..............................83
Table 3. 1 Data collection statistics .................................................................101
Table 3. 2 SOLVE statistics for SAD phasing using SeMet crystal .................102
Table 3. 3 Phasing statistics ...........................................................................103
Table 3. 4 Refinement statistics .....................................................................106
Table 3. 5 Major crystal contacts .................................................................115
LIST OF FIGURES

Fig. 1. 1 Flaviviruses classification. .................................................................20
Fig. 1. 2 Geographical distribution of dengue & Aedes aegypti. ....................21
Fig. 1. 3 Genome organization of the dengue genome .................................22
Fig. 1. 4 Flaviviruses replication cycle. ..........................................................25
Fig. 1. 5 Electron micrographs of flavivirus particles. ....................................25
Fig. 1. 6 Structure of dengue virus. .................................................................27
Fig. 1. 7 Illustration of protein expression and polyprotein processing. .........29
Fig. 1. 8 Membrane topology of the flavivirus structural proteins. ..............30
Fig. 1. 9 Dengue 2 Capsid proteins .................................................................32
Fig. 1. 10 Structure of the ectodomain of dengue E protein .........................35
Fig. 1. 11 Comparison of substrate specificities of NS3 proteases with their natural cleavage sites in viral polypeptide ....................................................41
Fig. 1. 12 Structures of dengue 2 NS3pro/NS2B and WNV NS3pro/NS2B with Bzl-Nle-Lys-Arg-Arg-H inhibitor .............................................................43
Fig. 1. 13 Schematic representation of dengue NS3 protein .........................44
Fig. 1. 14 Schematic representation of dengue 2 NS5 ....................................48
Fig. 1. 15 Crystal structures of dengue NS5 SAM and HCV NS5B RdRp .......50
Fig. 1. 16 DNA replication fork .................................................................52
Fig. 1. 17 Classification of helicases ...............................................................54
Fig. 1. 18 Topology diagrams of representative helicases ..............................62
Fig. 1. 19 Structures of PcrA complexes .........................................................64
Fig. 1. 20 Inchworm mechanism of PcrA helicase .........................................64
Fig. 1. 21 Structure of Rep helicase .............................................................66
Fig. 1. 22 Active rolling model for DNA unwinding and translocation by the Rep DNA helicase dimer. ................................................................. 67

Fig. 1. 23 Illustration of “descending molecular see-saw” mechanism of processive duplex unwinding by HCV NS3 helicase. ....................... 68

Fig. 1. 24 Proposed HCV NS3 helicase mechanism .................................. 70

Fig. 1. 25 Illustration of conserved motifs and structure of HCV NS3-dU8 helicase complex and YFV NS3-ADP complex. ........................... 72

Fig. 1. 26 Structure of SF3 helicases and SV40 hexamer. ......................... 74

Fig. 1. 27 The change of channel openings of SV40 helicase hexamer. ........ 75

Fig. 1. 28 Looping model of SV40 double hexamer unwinding. ................ 76

Fig. 1. 29 Structures of the T7 helicase. .................................................. 78

Fig. 1. 30 A four-site binding change model for T7 hexameric helicase. ........ 79

Fig. 2. 1 Cloning strategies. ..................................................................... 82

Fig. 2. 2 PCR product of DNS3H (1.3Kb). .................................................. 83

Fig. 2. 3 Confirmation of DNS3H with NcoI and HindIII .......................... 84

Fig. 2. 4 Overexpression of Trx-DNS3H. ................................................. 85

Fig. 2. 5 Flowchart of DNS3H purification protocol ................................. 86

Fig. 2. 6 IMAC purification of Trx-DNS3H. .............................................. 87

Fig. 2. 7 Anion exchange chromatography of Trx-DNS3H. ...................... 88

Fig. 2. 8 Optimization of enterokinase cleavage ..................................... 89

Fig. 2. 9 Removal of thioredoxin. ............................................................. 90

Fig. 2. 10 Gel filtration chromatography of DNS3H. ................................ 91

Fig. 2. 11 Overexpression of Selenomethionyl Trx-DNS3H ..................... 92

Fig. 2. 12 Purification of SeMet DNS3H. .................................................. 93

Fig. 2. 13 Mass spectrometry of native and SeMet DNS3H. ..................... 94
List of figures

Fig. 2. 14 Vector representation of relationship of (2) (3) and (4). .........................95
Fig. 2. 15 The phase doublet of SAD. .................................................................96
Fig. 2. 16 Schematic plot of phase distribution curve from phasing of SAD data. 97
Fig. 3. 1 DNS3H crystals. .........................................................................................99
Fig. 3. 2 Structure of DNS3H ..................................................................................107
Fig. 3. 3 Structural alignments of Flaviviridae helicase domains. .....................109
Fig. 3. 4 Stereo view of DNS3H and HCV helicase by superposition of the
Ca-carbon atoms of DNS3H (blue) and HCV helicase (red). ..............................111
Fig. 3. 5 Superposition of the Ca-carbon atoms of DNS3H (red) and YFV helicases
(blue). ......................................................................................................................112
Fig. 3. 6 Major crystal contacts. ................................................................................114
Fig. 3. 7 Superposition of the DNS3H Mn^2+–SO_4^2– (green sticks) and YFV helicase
ADP complex (yellow sticks)............................................................................117
Fig. 3. 8 DNS3H active site .........................................................................................118
Fig. 3. 9 Comparison of P-loops of six molecules in three crystals. ....................120
Fig. 3. 10 Electrostatic representation of DNS3H. ................................................122
Fig. 3. 11 Model of DNS3H-dU8 complex .............................................................124
Fig. 3. 12 Comparison of nucleic acid binding sites ............................................124
Fig. 3. 13 Schematic view of local rearrangements in motif V and neighboring
residues in Den and YFV helicases. .....................................................................126
Fig. 3. 14 Surface representation comparison of the nucleic acid binding site. ....127
Fig. 3. 15 Brownian motor mechanism of unidirectional translocation. .............130
Fig. 3. 16 Model of the dengue full length NS3 protein ........................................134
LIST OF ABBREVIATIONS (alphabetically)

2’OMTases: (nucleoside-2’-O-)-methyltransferase
Å: angstrom
AAV: adeno-associated virus
ADE: antibody-dependent enhancement
AdoMet: adenosyl-L-methionine
ADP: adenosine diphosphate
ADPNP: 5’-adenylyl-beta, gamma-imidodiphosphate
ADSC: area detector systems corporation
ARM: armadillo
ATP/ATPase: adenosine triphosphate / adenosine triphosphatase
BM: Brownian motor
CCP4: collaborative computational project number 4
CM: convoluted membrane
CNS: crystallography & NMR system
DC-SIGN: dendritic-cell-specific ICAM-grabbing non-integrin
DNS3H: dengue nonstructural protein 3 171-618
Den: dengue
DF: dengue fever
DHF: dengue hemorrhagic fever
ds: double-stranded
DSS: dengue shock syndrome
EDTA: ethylenediaminetetraacetic acid
EM: electron micrograph
ER: endoplasmic reticulum
GRP78/Bip: glucose-regulating protein 78
HCV: hepatitis C virus
HEL: helicase
IMAC: immobilized metal affinity chromatography
IPTG: isopropyl-β-D-1-thiogalactoside
IRES: internal ribosome entry site
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>KUN</td>
<td>kunjun</td>
</tr>
<tr>
<td>LB</td>
<td>luira broth</td>
</tr>
<tr>
<td>MAD</td>
<td>multiple-wavelength anomalous dispersion</td>
</tr>
<tr>
<td>MES</td>
<td>morpholineethanesulfonic acid</td>
</tr>
<tr>
<td>MTase</td>
<td>methyltransferase</td>
</tr>
<tr>
<td>NA</td>
<td>nucleic acid</td>
</tr>
<tr>
<td>NC</td>
<td>nucleocapsid core</td>
</tr>
<tr>
<td>NCRs</td>
<td>noncoding regions</td>
</tr>
<tr>
<td>ncs</td>
<td>non crystallographic symmetry</td>
</tr>
<tr>
<td>Ni-NTA</td>
<td>nickel-nitritriacetic acid</td>
</tr>
<tr>
<td>NITD</td>
<td>Novartis institute of tropical diseases</td>
</tr>
<tr>
<td>NLS</td>
<td>nuclear localization signal</td>
</tr>
<tr>
<td>nm</td>
<td>nanometer</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>NS</td>
<td>nonstructural</td>
</tr>
<tr>
<td>nt</td>
<td>nucleotides</td>
</tr>
<tr>
<td>NTP</td>
<td>nucleoside triphosphate</td>
</tr>
<tr>
<td>NTPase</td>
<td>nucleoside triphosphatase</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PROT</td>
<td>protease</td>
</tr>
<tr>
<td>r.m.s.</td>
<td>root mean square deviation</td>
</tr>
<tr>
<td>RdRp</td>
<td>RNA dependent RNA polymerase</td>
</tr>
<tr>
<td>RTPase</td>
<td>RNA triphosphatase</td>
</tr>
<tr>
<td>SAD</td>
<td>single-wavelength anomalous dispersion</td>
</tr>
<tr>
<td>SAM</td>
<td>S-adenosylmethionine dependent methyltransferase</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecyl sulphate-polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SeMet</td>
<td>selenomethionine</td>
</tr>
<tr>
<td>SF</td>
<td>superfamily</td>
</tr>
<tr>
<td>ss</td>
<td>single-stranded</td>
</tr>
<tr>
<td>SV40</td>
<td>simian virus 40</td>
</tr>
<tr>
<td>Trx</td>
<td>thioredoxin</td>
</tr>
<tr>
<td>Tris-HCl</td>
<td>Tris(hydroxymethyl)aminomethane hydrochloride</td>
</tr>
<tr>
<td>VP</td>
<td>vesicle packet</td>
</tr>
</tbody>
</table>
List of abbreviations

WNV: west nile virus

YF/YFV: yellow fever / yellow fever virus
CHAPTER 1

INTRODUCTION

1.1 Viruses

A virus is a microscopic parasite that infects cells in biological organisms. Viruses are obligate intracellular parasites; they can reproduce only by invading and controlling other cells as they lack the cellular machinery for self-reproduction. The term virus usually refers to those particles that infect eukaryotes, while the term bacteriophage or phage is used to describe those infecting prokaryotes. Typically these particles carry a small amount of nucleic acid (either DNA or RNA), surrounded by some form of protective coat consisting of proteins, lipids, glycoproteins or a combination. Importantly, viral genomes code not only for the proteins needed to package its genetic material, but for proteins needed by the virus during its life cycle.

The origins of viruses are not clear and there may not be a single evolutionary path that can account for all viruses. Some of the smaller viruses that have only a few genes may have originated from host organisms. Their genetic material could have been derived from transferrable elements like plasmids or transposons. Viruses with large genomes may represent extremely reduced microbes which established symbiotic relations with host organisms, allowing the loss of some genes needed for existence independent of a host (Flint et al., 2004).

The viral genome (either DNA or RNA) is encapsidated by a protective coat of protein called a capsid. The viral capsid can be spherical or helical or without
Chapter 1

symmetry and is composed of proteins encoded by the viral genome. In helical viruses, the capsid protein, or the nucleocapsid protein, binds directly to the viral genome. For example, in the case of the measles virus, one nucleocapsid protein binds every six bases of RNA to form a helix approximately 1.3 micrometers in length. This complex of protein and nucleic acid is called the nucleocapsid, and, in the case of the measles virus, is enclosed in a lipid "envelope" acquired from the host cell, in which virus-encoded glycoproteins are embedded. These are responsible for binding and entering the host cell at the start of a new infection.

Spherical virus capsids completely enclose the viral genome and do not generally bind as tightly to the nucleic acid as helical capsid proteins do. These structures can range in size from less than 20 nanometers up to 400 nanometers and are composed of viral proteins arranged in a icosahedral symmetry. The number of proteins required to form a spherical virus capsid is denoted by the "T-number", where 60T proteins are necessary. The complete virus particle is referred to as a virion. A virion is little more than a gene transporter, and components of the envelope and capsid provide the mechanism for injecting the viral genome into a host cell (Flint et al., 2004).

1.2 Classification of viruses

Viruses can be classified in several ways, such as by their geometry, by whether they have envelopes, by the identity of the host organism they can infect, by mode of transmission, or by the type of disease they cause. The most useful classification is probably by the type of nucleic acid the virus contains and its mode of replication, which was proposed by Nobel-prize winner David Baltimore.
The principle used in the Baltimore classification system is the so-called central dogma conceptualized by Francis Crick: DNA → RNA → Protein.

The various forms of viruses arise because one of the two strands of DNA in which all cellular life forms store their genetic information is redundant, so that viruses can have either single-stranded or double-stranded genomes. Furthermore, some viruses store their genome as RNA rather than as DNA. RNA arises in cells as an intermediate when genes are translated into proteins. RNA genomes of viruses can be encoded in two different directions: Either the genes are stored in the 5' to 3' direction (positive or + polarity), or the genes are stored in the opposite direction (negative or - polarity). According to Baltimore’s concepts, viruses can be classified into the following seven groups (Table 1.1):

<table>
<thead>
<tr>
<th>Class</th>
<th>Nucleic Acid</th>
<th>Examples</th>
<th>Envelope</th>
<th>Genome size (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>dsDNA</td>
<td>Poxvirus</td>
<td>Yes</td>
<td>130-375</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adenovirus</td>
<td>No</td>
<td>3.0-4.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Papillomavirus</td>
<td>No</td>
<td>5.3-8.0</td>
</tr>
<tr>
<td>II</td>
<td>ssDNA</td>
<td>Adeno-associated virus</td>
<td>No</td>
<td>5.0</td>
</tr>
<tr>
<td>III</td>
<td>dsRNA</td>
<td>Reovirus</td>
<td>No</td>
<td>18-31</td>
</tr>
<tr>
<td>IV</td>
<td>(+)ssRNA</td>
<td>Flaviviridae</td>
<td>Yes</td>
<td>10</td>
</tr>
<tr>
<td>V</td>
<td>(-)ssRNA</td>
<td>Coronavirus</td>
<td>Yes</td>
<td>16-21</td>
</tr>
<tr>
<td>VI</td>
<td>(reverse)RNA</td>
<td>HIV</td>
<td>Yes</td>
<td>9.7</td>
</tr>
<tr>
<td>VII</td>
<td>(reverse)DNA</td>
<td>Hepatitis B virus</td>
<td>Yes</td>
<td>3.1</td>
</tr>
</tbody>
</table>

*a Nucleic Acid type, ds, double-stranded; ss, single-stranded; (+), positive sense; (-), negative sense; reverse, reverse transcribing.*
1.3 *Flaviviridae*

*Flaviviridae* is a family of viruses that infect mammals and birds. It consists of three genera: the flavivirus, the pestiviruses and the hepaciviruses. In addition, a group of unassigned viruses, the GB agents, are waiting for formal classification within the family. The genome of the *Flaviviridae* viruses is a linear, single-stranded positive-sense RNA of about 10,000-11,000 nucleotides long. The 5′-terminus carries a methylated nucleotide cap or a genome-linked protein. The viral particle is enveloped and spherical, about 40-60 nm in diameter. Although they share similarities in virion morphology, genome organization and RNA replication strategy, members within this family exhibit diverse biological properties and lack of serologic cross-reactivity. The increasing significance of *Flaviviridae* as human and animal pathogens makes it relevant for detailed studies by academic community (Lindenbach and Rice, 2001).

1.4 Flaviviruses

The flaviviruses comprise a group of about 80 viruses many of which are anthropod-borne human pathogens (Lindenbach and Rice, 2001). They are widely distributed throughout the world, although a specific flavivirus may be geographically restricted to a continent or a particular region. Flaviviruses can cause various syndromes, ranging from benign febrile illnesses to severe diseases with hemorrhagic manifestations or major organ failure. Some neurotropic flaviviruses can produce severe destructive central nervous system disease with serious sequelae. The most well known flaviviruses are yellow fever, dengue, Japanese encephalitis and West Nile viruses (*Fig. 1.1*).
Fig. 1.1 Flaviviruses classification.

The relationships between selected flaviviruses are shown in the dendrogram on the left. The serological (serocomplex) and phylogenetic (clade and cluster) classifications of these flaviviruses are shown on the right (Adapted from Mukhopadhyay et al., 2005).

1.5 Dengue and dengue virus classification

Dengue is an endemic viral disease affecting tropical and subtropical regions around the world transmitted by *Aedes aegypti* mosquitoes. Dengue fever (DF) and its more serious forms, dengue hemorrhagic fever (DHF) and dengue shock syndrome (DSS), are becoming important public health problems. The global prevalence of dengue has grown dramatically in recent decades (Fig.1.2). The disease is now endemic in more than 100 countries in Africa, the Americas, the eastern Mediterranean, Southeast Asia, and the Western Pacific, threatening more than 2.5 billion people (Gubler, 1998). The World Health Organization estimates that there may be 50 million to 100 million cases of dengue virus infections worldwide every year, which result in 250,000 to 500,000 cases of DHF and...
24,000 deaths each year (Gibbons and Vaughn, 2002; World Health Organization, 1997).

The Dengue virus has been shown to have 4 subtypes (DENV-1 to DENV-4). These 4 subtypes are different strains of dengue virus that share 60-80% homology with each other. The dengue viruses genome is about 11 kilobases (kb) in length (Fig. 1.3). The RNA genome contains a type I 5’ cap (m7GpppAmpN2) and lacks a polyadenylate tail. Genomic RNA is translated using a single long open reading frame (ORF) as a large polyprotein. Surrounding the ORF are 5’ and 3’ noncoding regions (NCRs) of around 100 nucleotides (nt) and 400 to 700 nt, respectively (Lindenbach and Rice, 2001).

![Geographical distribution of dengue & Aedes aegypti.](image)

**Fig.1.2 Geographical distribution of dengue & Aedes aegypti.**

*Left panel: Geographical distribution of dengue (light shading) and dengue fever plus dengue haemorrhagic fever (dark shading); Right panel: Aedes aegypti taking blood meal.*
1.6 Pathogenesis of dengue virus

Dengue virus infection can lead to a range of symptoms, from benign dengue fever (DF) to severe dengue hemorrhagic fever (DHF) and dengue shock syndrome (DSS). Dengue fever (DF) is a severe, flu-like illness that affects infants, young children and adults, but seldom causes death. The clinical features of dengue fever vary according to the age and immune status of the patient. Infants and young children may have a non-specific febrile illness with rash. Older children and adults may have either a mild febrile syndrome or the classical incapacitating disease with abrupt onset and high fever, severe headache, pain behind the eyes, muscle and joint pains (Halstead, 1988; Henchal and Putnak, 1990; Halstead et al., 2005). The fever usually lasts five to seven days. A rash, typically macular or maculopapular and often confluent with the sparing of small islands of normal skin are observed. It usually appears near defervescence and maybe accompanied by scaling and pruritus (Jelinek et al. 1997; Schwartz et al. 1996). Other signs and symptoms include flushed faces, lymphadenopathy, inflamed pharynx, injected conjunctivae, mild respiratory and gastrointestinal symptoms. Patients with DF may have hemorrhagic manifestations, such as petechiae, purpura, or evidence of a positive tourniquet test for capillary fragility (Vaughn et al., 2000). Infection by

Fig. 1. Genome organization of the dengue genome.

Genes encoding structural and non-structural proteins are indicated by blue and green boxes, respectively. The 5' and 3' NCRs are indicated by red lines.
Chapter 1

dengue virus may also cause myocarditis, hepatitis (Lum et al., 1993), and neurological abnormalities, such as encephalopathy and neuropa-thies (Sumar-mo et al., 1978), but in very rare cases.

In some cases, possibly when successive infections by different serotypes occur, patients develop the much more severe forms of disease known as dengue hemorrhagic fever (DHF) and dengue shock syndrome (DSS), depending largely on the patient’s age and immunologic condition (Gibbons and Vaughn, 2002). Although the mechanisms for the development of DHF and DSS are not fully understood, the main risk factor for the development of DHF and DSS is thought to be secondary infection with another serotype (Guzman et al., 2002; Gubler, 1998). In this theory, cross-reactive but non-neutralizing anti-dengue antibodies elicited from the previous infection bind to the new infecting serotype and enhance viral uptake by monocytes and macrophages, a phenomenon known as “antibody-dependent enhancement” (ADE). ADE results in an amplified cascade of cytokines and complement activation, causing platelet destruction, endothelial dysfunction, and consumption of coagulation factors leading to plasma leakage and hemorrhagic manifestations (Guzman and Kouri, 2002; Halstead and O’Rourke, 1977; Halstead, 1979). In addition, the severity of the disease also depends on the viral strain, the age and genetic background (Guzman and Kouri, 2002; Rosen, 1977; Gubler, 1998). DHF is characterized by capillary leakage and hemostatic changes. Plasma leakage usually develops at the time of defervescence. Outset of DSS is often associated with some warning signs, such as abdominal pain, persistent vomiting, change in level of consciousness, sudden change from fever to hypothermia and decrease in platelet count (Gibbons and Vaughn, 2002).
1.7 Replication cycle of flaviviruses

The replication cycle of flaviviruses starts with (1) the attachment of the virion to the surface of host cell and subsequent entrance into cells by receptor-mediated endocytosis. Several primary receptors and low-affinity co-receptors for flaviviruses have been identified. (See 1.9.1.3) (2) Fusion of the viral and cell membranes is triggered by irreversible trimerization of the envelope (E) protein inside the acidic environment of endosome (Allison et al., 1995). (4) In the cytoplasm, the released single-strand, positive sense RNA ((+)-ssRNA) genome serves as mRNA to direct the translation of a polyprotein which is subsequently cleaved by both host and viral proteases. (3) Simultaneously, the (+)-ssRNA also serves as a template to direct the synthesis of progeny RNA genome on intracellular membranes. (5) Viral assembly occurs on the surface of the endoplasmic reticulum (ER). (6) These newly assembled viral particles, consisting of E, membrane (prM) proteins, lipid membrane, capsid and RNA genome, are not infectious, therefore immature, until the cleavage of prM by the host protease furin in the trans-Golgi network (Stadler et al., 1997; Elshuber et al., 2003). (7) Mature viruses, as well as subviral particles, those containing only glycoprotein and membrane but not capsid protein and RNA genome, are then released from the host cell by exocytosis (Fig. 1.4).
1.8 Structure of dengue virus

Flaviviruses are small, about 50 nm in diameter, spherical particles containing an electron-dense core of about 30 nm, surrounded by a lipid envelope (Murphy, 1980). Electron micrographs of virus particles (Fig.1. 5) reveal that the surface contains two viral proteins: E (envelope) and M (membrane) protein.

Fig.1. 5 Electron micrographs of flavivirus particles.

Negatively stained DEN2 virus particles isolated from an infected mouse brain homogenate via rate zonal sedimentation through a sucrose gradient. Black bar represent 100 nm. (Adapted from Lindenbach and Rice, 2003.)
Using cryo-electron microscopy (cyroEM) techniques, the structure of dengue virus particle was determined at a resolution of 24 Å (Kuhn et al., 2002). Image reconstruction revealed a spherical particle of 500 Å diameter with a smooth surface and icosahedral features (*Fig. 1. 6 A*). The external shells are thin, ordered and dense layers, corresponding to E and M glycoproteins. Beneath the protein shells are two concentric shells representing the lipid bilayer. Inside the lipid bilayer are the nucleocapsid and RNA genome which has a less ordered structure (*Fig.1. 6 B*) (Kuhn et al., 2002). By placement of tick-borne encephalitis E protein structure (Rey et al., 1995) into the density, the external surface of the dengue virus particle is revealed to be composed of 90 E dimers arranged in an unusual “herring-bone” manner with three monomers per icosahedral asymmetric unit (*Fig.1. 6 C*). The unusual tight packing suggests that rotational rearrangement around three-fold and five-fold axes maybe needed for E dimers to form fusogenic trimeric complexes (Kuhn et al., 2002). Like in Semliki Forest virus (Lescar et al., 2001), E protein trimerization is triggered at low pH during the conversion of E to its fusogenic form (Allison et al., 1995; Stiasny et al., 1996). The structures of immature dengue and yellow fever virus particles have also been determined to resolutions of 16 and 25 Å, respectively (Zhang et al., 2003). Unlike the mature form, the surface of the immature virus contains 60 icosahedrally-organized trimeric spikes, each consisting of 3 prM-E heterodimers (*Fig.1. 6 D*). In each prM-E heterodimer, E proteins protrude away from the viral membrane making its long axis tilt by an angle of 25° with respect to the viral surface. The fusion peptide on its outer tip is covered by the N-terminal portion of prM, consistent with the notion that prM protects the fusion peptide during maturation.
Fig. 1. 6 Structure of dengue virus.

(A) Dengue virus-2 cryoEM reconstruction at 24 Å resolution shown as a surface-shaded representation. Outline of one icosahedral asymmetric unit and the definition of the coordinate system. The scale bar represents 100 Å. (B) Cross-section showing the cryoEM electron density with a plot of the maximum (blue) and averaged (purple) electron density. Arrows indicate the position of the 5-fold and 3-fold icosahedral symmetry axes. Radial density sections shown at the defined radii, r1: nucleocapsid; r2: lipid bilayer; r3: M glycoprotein; r4: E
glycoprotein. Dark shading represents protein in higher density. Scale bar represents 175 Å. (C) Proposed low pH dependent rearrangement of E protein dimers into the fusogenic state. The icosahedral asymmetric unit is represented by the triangle, and 3 and 5 fold symmetry axes are illustrated. Small arrows indicate the proposed rotation of E protein. E protein domains I, II, III, and the fusion peptide are colored in red, yellow, blue and green, respectively (Kuhn et al., 2002).

(D) Structure of dengue immature particle. An icosahedral asymmetric unit is outlined in black. One of the 60 spikes is colored: prM is gray and E is green (Zhang et al., 2003). (Adapted from Kuhn et al., 2002; Zhang et al., 2003).

1.9 Dengue virus proteins

In infected cell, the released 5’ capped, single-strand, positive sense RNA genome serves as mRNA to direct the translation of viral proteins. Unlike flaviviruses, other members of Flaviviridae family, hepaciviruses and pestiviruses are not capped at the 5’ termini of the genome and instead utilize internal ribosome entry sites (IRES) to initiate translation (Poole et al., 1995; Lemon and Honda, 1997; Le et al., 1998; Rijnbrand and Lemon, 2000). Translation of the single long open reading frame produces a large polyprotein that is cleaved co- and post-translationally into at least 10 proteins (Fig. 1.7). This 374 kDa large polyprotein encodes both viral structural and non-structural proteins, with the N-terminal one-fourth encoding structural proteins (C-prM-E), followed by nonstructural proteins (NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5) (Rice et al., 1985). Cleavage of polyprotein chain into individual proteins involves both host- and viral-encoded proteases. In essence, host signal peptidase catalyzes the
cleavage among C-prM, prM-E, E-NS1, and near the N terminus of NS4B. A virus-encoded serine protease, NS3/2B protease complex, catalyzes the cleavage between NS2A-NS2B, NS2B-NS3, NS3-NS4A, NS4A-NS4B and NS4B-NS5 as well as internal sites within C, NS2A, and NS3. In addition, the furin protease from the host catalyzes the cleavage of prM protein into M protein, which is essential for virion particle maturation. However, the enzyme responsible for NS1-2A cleavage is currently unknown (Lindenbach and Rice, 2003).

![Diagram of viral protein expression and polyprotein processing](image)

**Fig 1.** Illustration of protein expression and polyprotein processing.

Viral serine protease cleavage sites are indicated by a downward arrow. Diamonds and a triangle indicate sites of cleavage with signal peptidase and furin enzyme, respectively. Processing of NS1-2A occurs by an unknown ER resident host enzyme, which is indicated by a question mark. Internal serine protease cleavage sites within NS2A and NS3 are also illustrated. In addition, the serine protease (PROT) and helicase (HEL) domains of NS3 and methyltransferase (MTase) and RdRP domains of NS5 are indicated. (Adapted from Lindenbach and Rice, 2003.)

1.9.1 Structural proteins

During polyprotein translation, the structural proteins traverse the ER membrane
several times by various signal sequences and membrane anchor domains (Fig. 1.8). Synthesis of capsid protein takes place at the cytoplasm side of the ER. The prM and E proteins are translocated into the lumen of the ER thanks to a hydrophobic signal sequence at the carboxyl terminus of capsid and prM protein, respectively. As a result, after proteolytic cleavage, the capsid protein and viral RNA are localized in the cytoplasm while the prM and E proteins remains in the lumen of ER and form a stable heterodimer (Lorenz et al., 2002; Konishi and Mason, 1993; Allison et al., 1995).

![Membrane topology of the flavivirus structural proteins](image)

Fig.1.8 Membrane topology of the flavivirus structural proteins.

The predicted orientation of the structural proteins across the endoplasmic reticulum (ER) membrane is shown. Transmembrane helices are indicated by cylinders, arrows indicate the sites of post-translational cleavage and the cleavage sites of specific enzymes are indicated by different colours. E, envelope; NS1, non-structural protein 1; prM, precursor to membrane protein. (Adapted from Mukhopadhyay et al., 2005.)

1.9.1.1 Capsid protein (C)

C protein is an 11 kDa, highly basic protein (Boege et al., 1983; Rice et al.,
Chapter 1

1985; Trent, 1977), responsible for encapsulation of viral RNA genome to form nucleocapsid core (NC). Charged residues, which mediate the RNA interaction, located at the N and C termini (Khromykh and Westaway, 1996), are separated by a short internal hydrophobic domain that mediates membrane association (Markoff et al., 1997). At the carboxyl terminus, a small hydrophobic sequence functions as a signal peptide to direct the translocation of prM into the lumen of ER. This sequence is cleaved by viral serine protease to form the mature C (Amberg et al., 1994; Lobigs, 1993; Yamshchikov and Compans, 1994).

The NMR structure of dengue 2 C protein (residues 21 to 100) shows that it forms a dimer in solution, with each monomer having four helices (α1-α4) connected by short loops (Fig.1. 9 A). The major dimer contact is established by extensive hydrophobic interaction between two pairs of antiparallel helices, α2-α2’ and α4-α4’. A striking feature of this structure is the distinctly asymmetric charge distribution on the surface, with the portion formed by α4-α4’ positively charged and the opposite face formed by α2-α2’ and α1-α1’ devoid of charge (Fig.1. 9 B). Based on the non-uniform charge distribution and the concave shape of the hydrophobic cleft, Ma et al., proposed that the α4-α4’ region interacts with RNA, whereas the apolar α2-α2’ region could interact with the viral membrane (Fig.1. 9 C) (Ma et al., 2004).

X-ray crystal studies of Kunjin capsid protein revealed a similar structure and confirmed a dimer formation of flavivirus C protein (Dokland et al., 2004). However, in contrast to dengue C protein, the concave hydrophobic surface formed by α2-α2’ is partially covered by α1 and α1’. Interestingly, molecules inside the
crystal stack in a manner of a tubule which is reminiscent of the long, filamentous structure commonly formed by HEAT and ARM repeat protein multimeric structures that are typically involved in protein-protein interactions (Dokland et al., 2004; Andrade et al., 2001).

![Fig. 1. 9 Dengue 2 Capsid proteins.](image)

**Fig. 1. 9 Dengue 2 Capsid proteins.**

(A) Dengue 2 C dimer structure, (B) Asymmetric charge distribution with blue positive charge and white apolar. Note: right panel has the same orientation as (A). Left panel is a view of right panel rotated by 180 degree. (C) Model for molecular interactions between structural components of flavivirus. The viral membrane is shown on top near the hydrophobic cleft, and the viral RNA is shown on the bottom near the positively charged surface. (Adapted from Ma et al., 2004)

1.9.1.2 Membrane protein (prM and M)

prM (26 kDa), is the glycoprotein precursor of M protein. It is translocated into the lumen of ER by a small hydrophobic signal peptide sequence at the C-terminal of C protein. There are two coordinated cleavage sites on either side of this
transmembrane signal peptide (Fig. 1. 8). The viral protease NS3/2B complex catalyzes the cleavage on the cytosolic side to produce the mature form of C and host signalase catalyze the cleavage on the luminal side to produce the N-terminus of prM. Interestingly, cleavage by signalase on the luminal side is not efficient until the C protein is removed (Amberg et al., 1994; Lobgis, 1993; Yamshchikov and Compans, 1994). Optimization in the signal sequence of YF that improves the signalase cleavage is lethal for virus production, suggesting that regulation of cleavage at this site plays a vital role other than the rate of production of prM (Lee et al., 2000). However, this lethal effect can be rescued by introducing second site mutations in the optimized signal sequence in revertant viruses (Lee et al., 2000).

The N-terminal region of prM contains one to three N-linked glycosylation sites (Chambers et al., 1990) and three disulfide links formed by six conserved cysteine residues (Nowak and Wengler, 1987). Inside the ER lumen, prM and E glycoproteins form a heterodimer rapidly after synthesis. Several evidences have proved that prM prompts the proper folding of E protein like a chaperone (Konishi and Mason, 1993; Lorenz et al., 2002) and protect the fusion peptide of E protein from undergoing rearrangement to the fusogenic form in the reduced pH environment of the early secretory pathway (Guirakhoo et al., 1991, 1992). Although the structure of a flavivirus prM/M has not yet been determined, recent studies in prM of JEV have shown that the C–terminus transmembrane regions and a single residue, His-99, were the most important region for interaction with E protein in JEV assembly and maturation (Lin and Wu, 2005). In the late state of flavivirus replication cycle, the pr fragment of prM is cleaved to M by host furin protease in trans-Golgi network and mature virions are released by exocytosis.
1.9.1.3 Envelope protein (E)

E, a 53 kDa glycosylated protein, is the major virion surface protein which mediates binding and membrane fusion. It belongs to type I membrane, class II fusion protein and contains adjacent transmembrane domains at the C-terminus that serve as anchor to the membrane and the signal sequence for NS1 translocation (Fig. 1.8) (Lindenbach and Rice, 2001). E protein has 12 highly conserved cysteine residues that form intramolecular disulfide bonds. The importance of these disulfide bonds to overall conformation and antibody-mediated virus neutralization has been revealed when reduced and denatured E protein as well as synthetic peptides failed to elicit virus-neutralizing antibodies in mice (Lindenbach and Rice, 2001; Roehrig et al., 2004; Halstead et al., 2005). Dengue E protein has two potential N-linked glycosylation sites, Asn-153 and Asn-67, and utilization of these sites varies with the four serotypes of dengue virus (Smith and Wright, 1985; Johnson et al., 1994). Implication of glycans on E protein in receptor binding has been found by the fact that dendritic-cell-specific ICAM-grabbing non-integrin (DC-SIGN), a mannose-specific, oligomeric C-type lectin on the cell surface, is essential for productive infection of dendritic cells (Navarro-Sanchez et al., 2003; Tassaneetrithep et al., 2003). Nevertheless, how receptor binding mediates the entry of dengue virus on the cell surface is still unclear. It is likely that the virus recognizes different receptors in human and mosquito cells (Hung et al., 2004). Heparan sulfate, a negatively charged glycan on the cell surface has been proposed to bind to the positively charged residues on the surface of domain III of E protein (Chen et al., 1997; Hung et al., 2004). However, heparan sulfate is so widely expressed that some more specific protein receptors, such as DC-SIGN,
GRP78/Bip (Glucose-regulating protein 78), and CD14-associated molecules should be required to target dengue virus to permissive cells types (Wu et al., 2000; Chen et al., 1999a; Jindadamrongwech et al., 2004; Tassaneetrithep et al., 2003).

X-ray crystallography studies of the ectodomain of dengue2 and 3 E proteins revealed an elongated head-to-tail dimer (Fig. 1. 10) that is oriented parallel to the viral membrane and has a slight curvature corresponding to the surface of the viral membrane. Each of the monomers has three distinct domains which is reminiscent of the E1 protein of alphaviruses (Lescar et al., 2001). Domain I consists of an 8-stranded central β-barrel that is flanked on one side by an elongated dimerization domain (domain II) containing the fusion peptide at its distal end, and by domain III on the other side. Domain III is an immunoglobulin (Ig)-like domain that is responsible for receptor binding. The fusion peptide of one E monomer is buried between domains I and III of the adjacent monomer within a dimer.

![Structure of the ectodomain of dengue E protein.](image)

**Fig. 1. 10 Structure of the ectodomain of dengue E protein.**

*Dimeric, pre-fusion conformation of dengue 2 E proteins (Modis et al., 2003). In one monomer, domain I, II and III are colored red, yellow and blue, respectively. Fusion peptide is shown in green. The other monomer is grey. (Adapted from Mukhopadhyay et al., 2005)*
1.9.2 Non-structural proteins

Except for three structural proteins, C, prM and E, the dengue viral genome encode seven non-structural proteins, NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5. NS1 and part of NS4B are translocated into the lumen of ER by hydrophobic signal sequences at the C-terminus of E protein and NS4A, respectively (Lindenbach and Rice, 2001). Polyprotein processing of the remaining five NS proteins takes place at the cytoplasm site of ER (Fig. 1.8). Non-structural proteins are widely involved in the early stage of virus life cycle.

1.9.2.1 NS1

NS1 is a 46 kDa glycosylated protein. It is translocated into the lumen of ER by a hydrophobic signal sequence localized at the C-terminus of the E protein which is cleaved cotranslationally by host signalase to generate the mature N-terminus of the protein (Chambers et al., 1990; Falgout et al., 1989; Falgout and Markoff, 1995). Proteolytic processing at the NS1/2A junction occurs by an unknown ER resident host enzyme (Falgout and Markoff, 1995). Flavivirus NS1 protein has two or three N-linked glycosylations sites and 6 disulfide bonds formed by 12 conserved cysteines (Lee et al., 1989; Mason, 1989). Arrangement of the disulfide linkage has been mapped to be C1/C2, C3/C4, C5/C6, C7/C12, C8/C10, and C9/C11 for dengue NS1 protein (Wallis et al., 2004).

In ER, immature NS1 exits as a hydrophilic monomer, then quickly associates to form a stable hydrophobic non-covalent homodimer which interacts with membranous components after maturation (Winkler et al., 1988; Winkler et al.,
Within infected cells, NS1 associates with intracellular organelles and serves as a cofactor in viral RNA replication (Mackenzie et al., 1996; Lindenbach and Rice, 1997; Muylaert et al., 1997). Alternatively, NS1 protein is exported along the secretory pathway to the cell membrane, where it remains anchored, possibly via a glycosylphosphatidylinositol group (Jacobs et al., 2000) or is released as a soluble hexamer (sNS1) from infected cells (Crooks et al., 1994; Flamand et al., 1999). Recent studies have shown that NS1 can be detected in the serum of dengue-infected patients with a correlation between NS1 expression and the development and severity of the disease (Alcon et al., 2002; Libraty et al., 2002; Young et al., 2000).

N-linked glycosylation of NS1 of dengue virus involves two conserved asparagines, Asn-130 and Asn-207, that are glycosylated with high mannose moieties. The resultant increase in hydrophobicity after glycosylation is important for dimer stability and membrane association (Winkler et al., 1988; Pryor and Wright, 1994). Interestingly, unlike the intracellular form of NS1, the secreted form comprises complex glycans on one site whereas the other remains in high mannose moieties. What functions are performed by the secreted form of NS1 is still unknown. However, evidence has shown that secreted NS1 plays an essential role in RNA replication (Lindenbach and Rice, 1997). Furthermore, secreted NS1 induces strong immunological responses and antibodies against cell surface-associated NS1 can direct the complement-mediated lysis of infected cells (Henchal et al., 1988; Schlesinger et al., 1985, 1993; Chung et al., 2006; Lin et al., 2005). The importance of NS1 to the flavivirus life cycle and immune response makes it an attractive target for development of antiviral therapeutics. Advances
have been made by using NS1 synthetic peptides or DNA vaccine against fatal flavivirus infections in experimental animals (Volpina et al., 2005; Wu et al., 2003).

1.9.2.2 NS2A/2B

NS2A is a small hydrophobic protein of about 22 kDa with unknown function (Lindenbach and Rice, 2001). As mentioned before, cleavage at the N-terminus NS1/NS2A junction is catalyzed by an unknown ER-resident host enzyme (Falgout and Markoff, 1995), whereas the C-terminus is generated by viral NS3/NS2B serine protease complex in the cytoplasm (Fig. 1.8), indicating that this protein must span the membrane. In addition, an alternative cleavage site upstream of the C-terminus of NS2A can be utilized by viral NS3/NS2B protease complex, leading to a truncated form, NS2Aα, which is about 2 kDa smaller (Chambers et al., 1990; Nestorowicz et al., 1994). Mutagenesis studies in YF indicated NS2Aα cleavage-site mutants (Lys190Ser) have unimpaired RNA replication producing virus particle devoid of the C proteins (Kummerer and Rice, 2002). This defect can be complemented in trans by supplying NS2A or NS2Aα and compensated for by revertants restoring the basic nature at the NS2Aα cleavage site, or by change at a second-site like Asp-343 (Glu-338 in dengue 2) to a short, uncharged side chain, in the helicase domain of NS3. Interestingly, NS2Aα was not necessarily produced in these restored mutants, indicating that the basic residue, Lys-190, plays an important role in the production of infectious particle. Furthermore, cryoimmunogold staining revealed that NS2A was localized to vesicle packets (VPs), the presumed sites of viral RNA replication (Mackenzie et al., 1998). In
vitro studies with a glutathione-S-transferase-KUN NS2A fusion protein showed that NS2A can bind to NS3, NS5, and the KUN 3’NCR (Mackenzie et al., 1998). Thus, the NS2A protein is likely to be involved in coordinating the transition between RNA packaging and RNA replication, two processes that have been shown to be functionally coupled (Khromykh et al., 2001).

NS2B is a 14 kDa small membrane-associated protein containing a conserved central hydrophilic region flanked by two hydrophobic domains which mediates the membrane association (Lindenbach and Rice, 2003). It forms a complex with the NS3 protease domain and acts as a cofactor necessary for activation of NS3 protease activity (Arias et al., 1993; Chambers et al., 1991, 1993; Falgout et al., 1991; Jan et al., 1995; Yusof et al., 2000). The minimum core sequence essential for NS3 protease activation has been mapped to 40 residues (Leu54 to Glu93 for dengue 2) within the central hydrophilic region (Chambers et al., 1993; Falgout et al., 1993; Leung et al., 2001). Mutations that disrupt NS3/NS2B interaction result in an inactive protease (Chambers et al., 1993; Clum et al., 1997; Droll et al., 2000; Falgout et al., 1993; Jan et al., 1995; Niyomrattanakit et al., 2004). However, deletion studies also revealed that without the hydrophobic flanking regions, the central hydrophilic region of NS2B provides only basal protease activity (Niyomrattanakit et al., 2004). Structures of dengue 2 and West nile virus (WNV) with their own minimum core region of NS2B have been determined recently and will be discussed later (Erbel et al., 2006).
1.9.2.3 NS3

Dengue NS3 protein is a large multifunctional protein of 68 kDa, endowed with protease, helicase, NTPase, as well as 5’-terminal RNA triphosphatase activities and plays an important role in viral polyprotein processing and genome replication (Lindenbach and Rice, 2001). The N-terminal 180 amino acids of NS3 comprises a serine protease domain with the protein NS2B acting as a membrane anchoring cofactor, necessary for proteolytic activity (Wengler, 1991; Chambers et al., 1993; Li et al., 1999, Yusof et al., 2000). The C-terminal domain is involved in viral RNA replication (Bartelma et al., 2002; Benaroch et al., 2004)

1.9.2.3.1 NS3 protease domain (NS3pro)

The N-terminal one-third of NS3 encodes a trypsin-like serine protease (Bazan and Fletterick, 1989; Gorbalenya et al., 1989a) with NS2B acting as a cofactor which is necessary for proteolytic activity (Wengler et al., 1991; Chambers et al., 1993; Li et al., 1999, Yusof et al., 2000). Like other serine proteases, NS3pro contains a catalytic triad, which for dengue 2 NS3 are residues His-51, Asp-75, and Ser-135 (Valle and Falgout, 1998). In infected cell, NS3pro/NS2B complex plays pivotal role in viral polyprotein processing. It mediates cleavage at NS2A/NS2B, NS2B/NS3, NS3/NS4A, NS4A/NS4B, NS4B/NS5 junctions and additional cleavage sites within C, NS2A, NS4A and NS3 itself (Lindenbach and Rice, 2003). Except for the NS2B/NS3 junction, which contains a glutamine residue at the P2 position, the NS3pro/NS2B complex preferentially cleaves a pair of dibasic amino acids (Arg or Lys) at the P1 and P2 positions, followed by an amino acid with a short side chain (Gly, Ala, or Ser) at the P1’ site (Lin et al., 1993) (Fig. 1.11).
Fig. 1. 11 Comparison of substrate specificities of NS3 proteases with their natural cleavage sites in viral polypeptide.

(A) Summary of the P4-P4' substrate specificities of the NS3 proteases. Single letter amino acid abbreviations are used above and below the schematic. (B) the P6-P6' boundary cleavage sites between NS2A/NS2B, NS2B/NS3, NS3/NS4A, and NS4B/NS5 are listed. Dengue 1, Singapore strain; Dengue 2, New Guinea C strain; Dengue 3, H87 strain; Dengue 4, Dominica strain. The P1 and P1' in each site is separated by a space to indicate the scissile bond. (Adapted from Li et al., 2005)

Several structures of NS3pro of *Flaviviridae* have been determined (Love et al., 1996; Kim et al., 1996; Murthy et al., 1999; Murthy et al., 2000; Erbel et al., 2006). Crystal structure of dengue 2 NS3pro/NS2B reveals a chymotrypsin-like fold with two β-barrels, each formed by six β-strands, with NS2B contributing an additional β-strand (residues 51-55 of NS2B) to the N-terminal β-barrel (Fig. 1. 12 A) (Erbel et al., 2006). Surprisingly, binding of NS2B render the NS3pro able to adopt a significant different conformation compared to the previously determined structure in the absence of NS2B (Murthy et al., 1999), with an r.m.s deviation of
2.1Å for 94 α-Carbons out of a total of 148 (Erbel et al., 2006). Comparatively, the difference between HCV NS3pro and NS3pro/NS4A is much smaller (Love et al., 1996; Kim et al., 1996). However, the previously solved structure was from a refolded protein with minimum protease activity, indicating that NS2B is crucial for stabilizing the catalytically active conformation of NS3pro (Murthy et al., 1999; Erbel et al., 2006). Deletion studies showed that even though the N-terminal part of NS2B is sufficient to stabilize the enzyme, as shown by NMR, the C-terminal part is important for activation and for binding the substrate (Erbel et al., 2006). In the structure of dengue 2 NS3pro/NS2B, the C-terminal part of NS2B is involved in crystal packing contacts (Erbel et al., 2006).

Crystal structure of WNV NS3pro/NS2B with the Bzl-Nle-Lys-Arg-Arg-H inhibitor gave new insight into the function of NS2B. In this structure, the C-terminal part of NS2B encircles the C-terminal β-barrel of NS3pro by making contact with strands B2a, B2b, and a loop between E1b and F1. Interestingly, involvement of C-terminal part of NS2B changed the conformation of C-terminal β-barrels of NS3pro resulting in reorientation of key residues for substrate or inhibitor recognition (Fig. 1.12 B) (Erbel et al., 2006). Notably, the structure revealed that residues Gly-83 to Gln-86 of NS2B also contributed to the formation of substrate S2 and S3 pockets, corroborating that the C-terminal part of NS2B is important for activation of NS3pro (Erbel et al., 2006).
Fig. 1. 12 Structures of dengue 2 NS3pro/NS2B and WNV NS3pro/NS2B with Bzl-Nle-Lys-Arg-Arg-H inhibitor.

NS3pro and NS2B are colored grey and yellow, respectively. Secondary structural elements are labeled. Residues from the catalytic triad are colored as grey sticks. Inhibitor is colored as orange sticks. Catalytic triad is colored as grey sticks. (Adapted from Erbel et al., 2006)

1.9.2.3.2 NS3 helicase domain (NS3hel)

The C-terminal region spanning residues 180 to 618 of the dengue 2 NS3 amino acid sequence comprises two motifs named Walker A, GK(S/T), and Walker B, DEx(D/H) (Fig. 1. 13). These motifs are present in a vast family of nucleotide binding proteins that participate in a wide variety of cellular functions by coupling NTP hydrolysis with directional movement, nucleic acid duplex destabilization, RNA processing, DNA recombination and repair (Walker et al., 1982; Singleton and Wigley, 2002). The presence of five additional conserved motifs places NS3 in superfamily 2 of RNA helicase/NTPases, according to the classification of
helicases into three major superfamilies (Gorbalenya and Koonin, 1993).

![NS3 protease and helicase domains present in NS3 are depicted as well as the conserved motifs in the amino-acid sequence.](image)

**Fig. 1.** Schematic representation of dengue NS3 protein.

**NS3 protease and helicase domains present in NS3 are depicted as well as the conserved motifs in the amino-acid sequence.**

1.9.2.3.2.1 NTPase/helicase activity

Functionally, the helicase and NTPase activities of the NS3 protein have been characterized for several members of the *Flaviviridae*, including hepatitis C virus (Gwack *et al.*, 1996), dengue virus (Li *et al.*, 1999; Benarroch *et al.*, 2004), West Nile virus (Borowski *et al.*, 2001), yellow fever virus (Warrener *et al.*, 1993), and Japanese encephalitis virus (Utama *et al.*, 2000). Dengue viruses and bovine viral diarrhea virus with impaired helicase activity are not able to replicate, demonstrating the importance of NS3 in the *Flaviviridae* life cycle (Grassmann *et al.*, 1999; Matusan *et al.*, 2001).

NTPase activity is involved in hydrolyzing the $\gamma$ phosphoric anhydride bond of NTP. It requires both Walker A and Walker B motifs in which the side chain of lysine of Walker A attacks the $\gamma$ phosphate of NTP and Walker B serves as a chelator for stabilization one divalent cation (Caruthers and Mckay, 2002). Mutations that disrupted the interaction between Walker A and NTP resulted in loss
of NTPase activity (Li et al., 1999; Matusan et al., 2001; Benarroch et al., 2004). NTPase activity is divalent cation (Mg$^{2+}$ or Mn$^{2+}$)-dependent (Li et al., 1999; Benarroch et al., 2004). Mutations disrupting coordination between magnesium and Walker B motif also abolished NTPase activity (Benarroch et al., 2004). Chelating agent such as EDTA also showed inhibition to NTPase activity (Li et al., 1999; Benarroch et al., 2004). NTPase activity of dengue NS3 helicase is very basal and can be stimulated by single-stranded homopolymeric RNA (Li et al., 1999). This RNA-stimulated NTPase activity is thought to be induced by conformation change within the protein upon RNA binding which results in a more kinetically favorable active site for NTP interaction (Preugschat et al., 1996). Mutagenesis and deletions studies have shown that a stretch of 21 residues (160 to 180) and a basic residues cluster$^{184}$RKRK are required for dengue NS3hel RNA stimulated NTPase activity (Li et al., 1999). Furthermore, recent studies showed that NS5 can stimulate NTPase activity in a dose-dependent manner up to the point of 1:1 molar ratio (Yon et al., 2005).

Helicase activity involves separation of duplex oligonucleotides into single strands. Like the HCV helicase, dengue NS3hel has a 3’ to 5’ directionality and can unwind both RNA and DNA duplexes (Gwack et al., 1996; Paolini et al., 2000; Tai et al., 1996; Li et al., 1999; Benarroch et al., 2004). It is normally thought that helicase utilizes the energy released from NTP hydrolysis to unwind DNA or RNA duplexes as shown by the existence of mutants devoid of NTPase activity which showed a loss of helicase activity (Li et al., 1999; Benarroch et al., 2004). However, several studies have indicated that the NTPase and helicase activities can be functionally uncoupled (Borowski et al., 2001; Gu et al., 2000; Heilek and
Chapter 1

Peterson, 1997; Hsu et al., 1998; Matusan et al., 2001; Paolini et al., 2000). For example, a Met-283 to Phe mutation in dengue NS3hel (residues 161 to 618) showed reduced ATPase and increased helicase activity (Matusan et al., 2001). Notably, dengue NS3hel is less efficient in both NTPase and helicase activity compared to the full length NS3 protein (Yon et al., 2005).

1.9.2.3.2.2 RTPase activity

In addition to its NTPase and helicase activity, Flavivirus NS3hels also possess RNA triphosphatase (RTPase) activity (Bartelma and Padmanabhan, 2002; Lindenbach and Rice, 2003; Benarroch et al., 2004). RTPase activity involves the hydrolysis of the $\gamma$ phosphoric anhydride bond of triphosphorylated RNA which is the first of the three sequential enzymatic reactions for addition of 5’ cap to RNA. Mutagenesis studies and competition experiments have concluded that RTPase and NTPase activities shared a common active site in Flavivirus NS3 (Bartelma and Padmanabhan, 2002; Benarroch et al., 2004). However, Bartelma and Padmanabhan reported that RTPase activity of dengue 2 NS3 is divalent-cation-independent for full-length NS3 (Bartelma and Padmanabhan, 2002) which is similar to the observation made for a proteolytic fragment of WN NS3 (Wengler, 1993). Interestingly, Benarroch et al., reported that RTPase activity of dengue 2 NS3 is divalent-cation-dependent because mutation destroying the coordination between magnesium and Walker B motif abrogated both RTPase and NTPase activity for a truncated form of NS3 (Benarroch et al., 2004). Remarkably, divalent cation dependence of RTPase activity proposed by Benarroch et al., is more plausible because it solves the contradiction of two Mg$^{2+}$-dependent activities and one Mg$^{2+}$-independent activity sharing the same active site.
(Benarroch et al., 2004). Furthermore, recent studies showed that NS5 can stimulate RTPase activity in a dose-dependent manner up to the point of 1:1 molar ratio (Yon et al., 2005).

1.9.2.4 NS4A/4B

NS4A and NS4B are small hydrophobic proteins with molecular weights of 16 kDa and 27 kDa, respectively (Lindenbach and Rice, 2003). NS4A colocalizes to VPs, membrane structures presumed to be the sites of RNA replication, suggesting an involvement in this function (Lindenbach and Rice, 1999). Furthermore, NS4A was also found in convoluted membranes (CMs), the site where polyprotein is processed (Mackenzie et al., 1998). The C-terminal end of NS4A acts as a signal peptide for the translocation of NS4B into the ER lumen (Lindenbach and Rice, 1999). Interestingly, cleavage at the NS4A/NS4B junction by signal peptidase requires cleavage at a site just upstream of the signal peptide by the viral serine protease (Lin et al., 1993; Preugschat and Strauss, 1991). In addition, unprocessed NS3/4A and NS4A/NS4B forms have been observed (Chambers et al., 1990; Lobigs, 1992; Preugschat and Strauss, 1991). NS4B is a transmembrane protein that localizes to the replication sites and to the nucleus (Westaway et al., 1997). The structure and function of NS4A and NS4B remain to be determined.

1.9.2.5 NS5

The dengue NS5 protein is the largest multifunctional protein of 103 kDa encoded by the viral genome (Lindenbach and Rice, 2003). It contains two
separate functional domains (Fig. 1. 14). The N-terminal domain one third of
NS5 is an S-adenosylmethionine-dependent methyltransferase involved in capping
viral RNA, as it contains two highly conserved motifs within several groups of
methyltransferases from a wide range of species (Koonin, 1993). The C-terminal
domain possesses RNA-dependent-RNA-polymerase (RdRp) activity, which is
essential for viral replication (Ackermann & Padmanabhan, 2001; Guyatt et al.,
2001; Khromykh et al., 1998; Steffens et al., 1999; Tan et al., 1996). The
interdomain region composed of residues 320 to 405 contains two nuclear
localization signals (NLS), β1 NLS and α/β NLS, which play important role for
trafficking NS5 into the nucleus as well as for directing the interaction between
NS5 and NS3 (Forwood et al., 1999; Johansson et al., 2001; Brooks et al., 2002).

Fig. 1. 14 Schematic representation of dengue 2 NS5.

The S-adenosylmethionine dependent methyltransferase (SAM) domain (Egloff
et al., 2002), β1 NLS (Brooks et al., 2002), α/β NLS (Brooks et al., 2002; Forwood
et al., 1999), and RNA-dependent-RNA-polymerase (Koonin, 1993) domain are
illustrated as cyan, blue, yellow and red bars, respectively. The
S-adenosyl-L-methionine (AdoMet) binding motif (Koonin, 1993) and the
conserved polymerase motifs A to D (Poch et al., 1989) are shown in black.
Asterisk indicates the end of the characterized SAM domain (Egloff et al., 2002).

SAM activity is the last of the three sequential enzymatic reactions for addition
of a 5’ cap to RNA. Cap structure of many viral and eukaryotic mRNAs is essential for stability of the RNAs, efficient binding to ribosomes and initiation of translation (Filipowicz, 1978; Furuichi and Shatkin, 2000). The cap structure is normally synthesized in the nucleus, thus viruses replicating in the cytoplasm normally encode their own enzymes for capping mRNAs (Ahola and Ahlquist, 1999; Ahola and Kaariainen, 1995; Chen et al., 1999b; Magden et al., 2001; Ramadevi et al., 1998). Unlike Flaviviruses, Pestiviruses and Hepaciviruses do not have SAM domain in their NS5 proteins, therefore they do not possess the 5’ cap structure in their viral genomes. Instead, they utilize an internal ribosome entry site (IRES) to initiate translation (van Regenmortel et al., 2000).

The crystal structure of the dengue NS5 SAM domain has been determined (Egloff et al., 2002; Benarroch et al., 2004). Structurally, it adopts an overall globular fold with three subdomains. The core subdomain 2 (in yellow) has a fold similar to the catalytic domain of other AdoMet-dependedent MTases (Fauman et al. 1999) flanked by two extensions at either side (Fig.1. 15 A) (Egloff et al., 2002). Interestingly, S-adenosyl-homocysteine (AdoHcy) which originates from E. coli cells was copurified and identified to bind to a central cleft between β-strands 1 and 4, where AdoMet and the substrate bind and methyltransfer occurs (Fauman et al., 1999). Structure conservation of active site residues (KDKE) between NS5 SAM domain and other 2’OMTases suggests that dengue NS5 SAM domain is a (nucleoside-2’-O-)-methyltransferase (Egloff et al., 2002).
Fig. 1.15 Crystal structures of dengue NS5 SAM and HCV NS5B RdRp.

(A) Dengue NS5 SAM domain. The three subdomains are colored in red, yellow and blue, respectively. Secondary structural elements are labelled. AdoHcy is colored as yellow sticks. (B) HCV NS5B RdRp. Subdomains (Fingers, Palm and Thumb) are labeled. Motifs A-F are colored as red, green, yellow, light purple, dark purple, and black, respectively. The long β sheet (priming loop) of thumb domain is colored in orange (Adapted from Lesburg et al., 1999; Egloff et al., 2002)

The C-terminal two-thirds of NS5 (Fig. 1.14) possesses RNA-dependent-RNA polymerase (RdRp) activity, which has been widely studied for several flaviviruses (Ackermann and Padmanabhan, 2001; Guyatt et al., 2001; Khromykh et al., 1998; Steffens et al., 1999; Tan et al., 1996). Sequence comparison revealed that this region contains four conserved motifs (A to D) that are found in all polymerases and two additional conserved motifs (E and F) that are characteristic of positive-strand RNA viruses only (Koonin, 1991; Rice et al., 1985). The importance of RdRp as the central enzyme in replication of the flaviviruses genome suggests that it is an ideal target for antiviral drug design.
Crystallographic studies of HCV NS5B revealed a compact globular, U-shaped structure (Ago et al., 1999; Bressanelli et al., 1999; Lesburg et al., 1999). Like all known polymerase structures, HCV NS5B contains the fingers, palm, and thumb domains but with many notable differences (Ago et al., 1999; Bressanelli et al., 1999; Lesburg et al., 1999). For instance, in the thumb domain of HCV NS5B, a long double-stranded antiparallel β sheet (orange) extends down into the active site cleft and occludes the putative dsRNA binding site. This long β sheet might play a role in maintaining the stability of binding to the RNA duplex or sequence-specific recognition or initiation from the 3’ terminus of HCV genomic RNA (Fig. 1.15 B).

1.10 Helicases: An ATP-fueled molecular motor

Molecular motors are biological "nanomachines" that are essential agents of movement in living organisms. A motor is defined as a device that consumes energy in one form and converts it into motion or mechanical power. Protein-based molecular motors, or protein motors, convert the chemical energy present in ATP into mechanical energy. Protein motors are widely involved in metabolism of cells, such as myosin responsible for muscle contraction, kinesin moving cargo within the cell along microtubules, RNA polymerase transcribing RNA from a DNA template, topoisomerases reducing supercoiling of DNA in the cell and so on.

Helicases are a class of molecular motor proteins vital to all living organisms. They use the energy of nucleoside 5’-triphosphate (NTP) to unidirectionally translocate along NA and unwind the complementary strands of the NA duplex. Helicases can also destabilize the secondary structure of RNA, remove NA associated proteins and thread NA through various pores. The universal presence
of helicases in prokaryotes, eukaryotes and virus, reflects their fundamental importance in DNA and RNA metabolic processes, including DNA replication (Fig. 1.6), repair, recombination, transcription, ribosome biogenesis, translation, RNA splicing, RNA editing, RNA transport, RNA degradation, bacterial conjugation, and viral packaging/unpackaging (de la Cruz et al., 1999; Lüking et al., 1998; Matson, 1991; Hall and Matson, 1999). Consistent with their vital roles in NA metabolism, helicases have been implicated in a variety of human genetic disorders, such as Werner syndrome, Bloom syndrome and xeroderma pigmentosum (Ellis, 1997).

**Fig. 1.6 DNA replication fork.**

The two ssDNA stands unwound by DNA helicases serve as templates for the DNA polymerase to synthesize new, complementary DNA (leading and lagging strand).

### 1.10.1 Helicases classification

All helicases share several common biochemical properties, including NA binding, NTP binding and hydrolysis, and NTP-driven unidirectional unwinding of NA duplex (Caruthers and McKay, 2002). According to their substrate or directionality, helicases can be divided into different types: DNA helicases and RNA helicases or 3’ to 5’ and 5’ to 3’ helicases. Furthermore, based on the active assembly state, helicases can be grouped as monomeric or multimeric helicases. In general, monomeric helicases require two different NA binding sites, one for ssNA
Chapter 1

and the other for dsNA. PcrA (Velankar et al., 1999) and bacteriophage T4 Dda helicase (Morris et al., 2001) are the few examples of monomeric helicases. In contrast, multimeric helicases, such as SV40 (hexameric), require only one NA binding site per monomer. The feature of a multimeric helicase is that it possesses multiple NA binding sites, enabling it to bind both ss and dsNA.

In 1993, Gorbalenya and Koonin recognized that some distantly related helicases share short, conserved amino acid motifs. Discriminated both by the number of distinct motifs that are identified within each group and by differences in the consensus sequences for motifs that are shared by more than one group, they proposed a general classification of helicases into three major groups, SF1 to 3 and two additional families, F4 and F5 (Gorbalenya and Koonin, 1993) (Fig. 1.17).
Fig. 1. Classification of helicases.

(A) Circles represent the groups of helicases, with their diameters roughly proportional to the number of proteins in each group. The distance between the circles approximately reflects the relationship between respective groups of helicases. Small circles encompassed by larger circles of three large superfamilies (SF1-SF3) represent distinct groups of helicases. Only selected groups, namely those including proteins with demonstrated helicase activity and/or containing numerous members, are labeled. Empty circles within SF1 and SF2 represent those groups including a single putative helicase. Generally, groups are designated after the prototype (the best characterized) protein. Rho helicase is related not to other helicases but to proton-translocating ATPases (Adapted from Gorbalenya and Koonin, 1993). (B) Helicases are represented as a bar with the N-terminal on the left and C-terminal on the right. Open boxes represent the conserved helicase motifs and letters inside the boxes are the consensus amino acid sequences of each motif. Labels above the open boxes are the names assigned to the motifs. Walker A and B motifs of SF1 are also indicated. The consensus amino acid sequence of SF1 and...
SF2 were taken from Gorbalenya et al., (1989b). Single-letter amino acid abbreviations indicate the presence of an amino acid in more than 75% of the family members. The consensus sequences for SF3 motifs were taken from Gorbalenya et al., (1989b). Upper-case amino acid abbreviations for SF3 indicate the presence of the residue in more than 50% of all family members and lower-case abbreviations indicate the presence of the residue in more than 50% of either the DNA viral or RNA viral family members. The consensus sequences of F4 motifs were taken from Ilyina et al., (1992). Single-letter amino acid abbreviations were used when a residue was present in at least six of the seven consensus sequences. In the consensus sequences, ‘+’, ‘o’, ‘x’ represents hydrophobic residue, hydrophilic residue and any residue, respectively (Adapted from Hall and Matson, 1999).

1.10.1.1 SF1 & SF2 superfamilies

SF1 and SF2 are the two largest and most closely related families of helicases (Gorbalenya and Koonin, 1993). They contain a large number of DNA and RNA helicases from archaea, eubacteria, eukaryotes and viruses, including representatives that unwind duplexes in a 3’ to 5’ or 5’ to 3’ direction (Gorbalenya and Koonin, 1993; Caruthers and McKay, 2002). Sequence analysis revealed that members of SF1 and SF2 share at least seven conserved motifs (I, Ia, II-VI) (Hall and Matson, 1999), which are usually clustered in a core region of 200 to 700 amino acids (Tuteja and Tuteja, 2004) (Fig.1. 17 B, SF1 and SF2). More recently, the existence of a new motif, named Q motif, which is upstream of motif I, was proposed in the DEAD-box family of helicases (Tanner et al., 2003).
1.10.1.1 Helicase motifs

Motifs I and II, which are usually referred to as Walker A and B motifs (Walker et al., 1982), are the most conserved motifs across all of the helicases families (Gorbalenya and Koonin, 1993). They include residues that interact with MgATP or MgADP. The Walker A motif, or phosphate binding loop (P-loop), which was typically defined as having a GxxxxGKT consensus (Walker et al., 1982), minimally requires the three final residues GK(T/S). It has been shown that the amino group of the lysine side chain of GK(T/S) interacts with the $\beta$ and $\gamma$ phosphates of ATP. The hydroxyl group of the threonine or serine residue ligates the Mg$^{2+}$ ion (Caruthers and McKay, 2002). Mutations of lysine or the last threonine abolished ATPase activity (Rozen et al., 1989; Blum et al., 1992; Pugh et al., 1999; Caruthers and McKay, 2002; Tanner et al., 2003; Cordin et al., 2004).

The Walker B motif, which was originally defined as a single aspartic acid residue (Walker et al., 1982), takes the general form DExx across the SF1 and SF2 superfamilies (Gorbalenya and Koonin, 1993). The carboxyl group of the aspartic acid coordinates the Mg$^{2+}$ ion of MgATP or MgADP, whereas the glutamic acid is suggested to act as a catalytic base in ATP hydrolysis (Caruthers and McKay, 2002). Other residues within Walker B motif of SF1 and SF2 helicases are proposed to mediate interactions with other motifs in a family-specific manner. In SF2 helicases, the Walker B motif has the form of DEx(D/H). In some crystal structures, such as HCV NS3 (Yao et al., 1997; Cho et al., 1998; Kim et al., 1998; Yao et al., 1999) and UvrB (Machius et al., 1999; Nakagawa et al., 1999; Theis et al., 1999), the fourth residue, histidine, interacts with a conserved glutamine of motif VI. In contrast, no direct interaction between the equivalent residues of motif
II and VI has been observed structurally in the DEA(D/H) RNA helicases. However, a covariation in sequence, with the correlation D (motif II) ↔ H (motif VI) or H (motif II) ↔ Q (motif VI), is observed, suggesting a direct interaction between these two residues (Caruthers and McKay, 2002).

Structural studies of several SF1 and SF2 helicases bound with DNA substrates (Korolev et al., 1997; Velankar et al., 1999; Kim et al., 1998) revealed that motifs Ia and IV contribute specific interactions with oligonucleotides. In addition, another loop region with the sequence TxGx, which is currently referred to as Ib motif, is also seen to interact with ssDNA (Caruthers and McKay, 2002). Mutants targeting motif Ia of DEAD-box helicases affect both ATPase and helicase activities (Svitkin et al., 2001). However, similar mutations in motif Ia of the DEAH-box helicases have a minor or no effect on yeast growth (Schneider et al., 2004). Furthermore, a single mutation (Val-539 to Ile) in motif Ia of Prp22 can salvage helicase activity from a defect in motif III in vitro, presumably by facilitating the opening of the interdomain cleft upon ATP hydrolysis. All these observations suggest that motif Ia contributes not only to binding of oligonucleotides, but also to structural rearrangements upon ATP binding and hydrolysis (Cordin et al., 2006).

The functions of motifs III and V are currently not completely understood. In addition, motifs III and V vary significantly in both length and amino acid sequence across different helicase families (Caruthers and McKay, 2002). Structure analysis suggested that they might be involved in a variety of interactions, including ligation of MgATP or MgADP, formation of specific salt bridges or hydrogen bonds between domains, binding to oligonucleotides and coupling of
ATPase to helicase activity (Velankar et al., 1999; Subramanya et al., 1996; Soultanas et al., 1999).

Motif VI is located at the interface close to the NTP binding pocket. The consensus sequence shared by both SF1 and SF2 is an arginine in the middle of the motif (Gorbalenya and Koonin, 1993). In the structures of PcrA complexed with AMPPNP, the guanidinium group of this arginine (Arg-610) makes direct contact with the \( \gamma \) phosphate of the nucleotide (Velankar et al., 1999; Soultanas et al., 1999). However, in the structure of nucleotide-free SF2 helicases, such as HCV helicase (Yao et al., 1997; Cho et al., 1998; Kim et al., 1998; Yao et al., 1999), UvrB (Machius et al., 1999; Nakagawa et al., 1999; Theis et al., 1999) and the DEAD-box helicases (Caruthers et al., 2000; Story and Abelson, 2001), the guanidinium groups of the equivalent arginines are away from their respective nucleotide binding site. Nevertheless, owing to the nature of interdomain flexibility of helicases, a small local rearrangement is sufficient to orient the side-chain of arginine to the proper position in order to interact with the \( \gamma \) phosphate of the nucleotide.

1.10.1.2 SF3 superfamily

The SF3 superfamily consists of helicases encoded only by small DNA and RNA viruses. Sequence analysis revealed that members of SF3 contain three conserved motifs (Gorbalenya and Koonin, 1993), Walker A and B motifs and motif C, within a limited \(~100\) amino acid region (Fig. 1.17 B, SF3). Later, another
conserved motif, named B’, located between motifs B and C was identified (Koonin, 1993).

1.10.1.3 Family 4 & 5

Family 4 is a small group of helicases that are found in bacterial and bacteriophage, consisting of proteins that are related in sequence to the *Escherichia coli* DnaB protein. They generally form a hexamer structure and are invariably associated physically with DNA primases (Ilyina *et al.*, 1992) to unwind DNA in a 5’ to 3’ direction (Caruthers and McKay, 2002). Helicases of this family contain five conserved motifs, including the Walker A, B motifs and other motifs that participate in DNA binding or ATP binding and hydrolysis (Caruthers and McKay, 2002) (*Fig.1. 17 B, F4*). Finally, an additional group, family 5, exemplified by the transcription termination factor Rho, was recognized as a family with sequence similarity to the β subunit of proton-translocating ATPases (Caruthers and McKay, 2002).

1.10.2 Helicase structures and mechanisms

Several crystal structures from different helicase superfamilies have been reported, including the DNA helicases PcrA from *Bacillus stearothermophilus* (Subramanya *et al.*, 1996) and Rep from *Escherichia coli* (Korolev *et al.*, 1997) that are representatives of SF1, as well as the RNA helicase from HCV (Yao *et al.*, 2002).
1997; Cho et al., 1998; Kim et al., 1998) and UvrB (Nakagawa et al., 1999), which are members of SF2. Representatives of each family are displayed in Table 1. 2.

A “core” α/β structure motif of about 150 amino acids is conserved across all classes of helicases (except family 5) and bears structural similarity with the RecA protein, a protein involved in homologous DNA recombination (Story and Steitz, 1992) (Fig. 1. 18). In particular, for SF1 and SF2 helicases, this “RecA-like” core is visible as a tandem of parallel structures that has probably arisen through gene duplication. Conversely, SF3 and F4 helicases only have one “RecA-like” core (Caruthers and McKay, 2002) (Fig. 1. 18).
### Table 1. 2 Representatives of helicases

<table>
<thead>
<tr>
<th>Protein</th>
<th>Group</th>
<th>PDB ID</th>
<th>Ligand</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>PcrA</td>
<td>SF1</td>
<td>1PJR</td>
<td>ADP</td>
<td>Subramanya et al., 1996</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1QHH</td>
<td>ADPNP</td>
<td>Soultanas et al., 1999</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2PJR</td>
<td>Product</td>
<td>Velankar et al., 1999</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3PJR</td>
<td>Substrate</td>
<td>Velankar et al., 1999</td>
</tr>
<tr>
<td>Rep</td>
<td>SF1</td>
<td>1UAA</td>
<td>ADP/ssDNA</td>
<td>Korolev et al., 1997</td>
</tr>
<tr>
<td>HCV</td>
<td>SF2</td>
<td>1HEI</td>
<td>-</td>
<td>Yao et al., 1997</td>
</tr>
<tr>
<td>NS3</td>
<td></td>
<td>1A1V</td>
<td>-</td>
<td>Kim et al., 1998</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8OHM</td>
<td>ssDNA</td>
<td>Cho et al., 1998</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2F55</td>
<td>ssDNA</td>
<td>Mackintosh et al., 2006</td>
</tr>
<tr>
<td>UvrB</td>
<td>SF2</td>
<td>1C4O</td>
<td>-</td>
<td>Machius et al., 1999</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1D2M</td>
<td>-</td>
<td>Nakagawa et al., 1999</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1D9X</td>
<td>-</td>
<td>Theis et al., 1999</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1D9Z</td>
<td>ATP</td>
<td>Theis et al., 1999</td>
</tr>
<tr>
<td>YFV</td>
<td>SF2</td>
<td>1YKS</td>
<td></td>
<td>Wu et al., 2005</td>
</tr>
<tr>
<td>NS3</td>
<td></td>
<td>1YMF</td>
<td>ADP</td>
<td>Wu et al., 2005</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1N25</td>
<td>-</td>
<td>Li et al., 2003</td>
</tr>
<tr>
<td>SV40</td>
<td>SF3</td>
<td>1SVO</td>
<td>-</td>
<td>Gai et al., 2004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1SVM</td>
<td>ATP</td>
<td>Gai et al., 2004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1SVL</td>
<td>ADP</td>
<td>Gai et al., 2004</td>
</tr>
<tr>
<td>T7gp4</td>
<td>F4</td>
<td>1CR0</td>
<td>-</td>
<td>Sawaya et al., 1999</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1CR2</td>
<td>dATP</td>
<td>Sawaya et al., 1999</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1E0J</td>
<td>ADPNP</td>
<td>Singleton et al., 2000</td>
</tr>
<tr>
<td>Rho</td>
<td>F5</td>
<td>1A62</td>
<td>-</td>
<td>Allison et al., 1998</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2A8V</td>
<td>RNA</td>
<td>Bogden et al., 1999</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1PV4</td>
<td>ssDNA</td>
<td>Skordalakes and Berger, 2003</td>
</tr>
</tbody>
</table>
Chapter 1

Fig. 1. 18 Topology diagrams of representative helicases.

Yellow, conserved “RecA-related” ‘core’. Red, variable structural elements in domains with a “RecA-like” core. Green and blue, additional structural domains. The schematic on the upper right summarizes the positions in the topology of the “RecA-like” core of the seven conserved motifs defined by (Gorbalenya and
1.10.2.1 Structures and mechanisms of SF1 helicases

The first helicase structure solved was PcrA, a protein of Bacillus stearothermophilus, which belongs to the SF1 (Subramanya et al., 1996). PcrA is composed of four domains, two parallel “RecA-like” cores, 1a and 2a, which comprise all the conserved motifs, and two additional α-helical domains, 1b and 2b, which form a single insertion within the polypeptide sequence of the two “RecA-like” domains. The bound ADP is located at the bottom of the cleft between subdomains 1a and 2a (Subramanya et al., 1996). In 1999, two structures of PcrA helicase in complex with DNA substrate were solved (Velankar et al., 1999). In the first structure, PcrA is bound to a partial duplex DNA (10 base pairs and a 7 base 3' single stranded overhang) and ADPNP, a non-hydrolysable ATP analogue. This structure represents a snapshot of the helicase before ATP hydrolysis and is therefore referred to as the "substrate complex". The second structure consists of PcrA bound to the same DNA substrate and a sulphate ion located at the position of γ phosphate after ATP hydrolysis, thereby mimicking a "product complex". In both complexes, the protein is monomeric. Comparison of two complexes revealed significant ligand-induced conformational changes that are important for the catalytic mechanism of the enzyme, including closure of the cleft between domains 1a and 2a, and significant movements of domains 1b and 2b that alter their positions relative to each other (Fig.1. 19). The complexes provide evidence consistent with an “inchworm” mechanism (Fig.1. 19).
Fig. 1. 19 Structures of PcrA complexes.

(A) The product complex and (B) the substrate complex. Domains are colored in green, yellow, red, and blue, respectively. The bound DNA is colored in magenta, ADPNP and sulphate ion in gold (Adapted from Velankar et al., 1999).

In the inchworm model, the enzyme monomer is bound to ssDNA and then translocates along the DNA strand to the fork region, probably upon binding ATP. Helix destabilization and release of one of the ssDNA strands takes place as ATP is hydrolyzed (Velankar et al., 1999) (Fig. 1. 20).

---

Fig. 1. 20 Inchworm mechanism of PcrA helicase.
(a) A model for helicase activity with regard to the large conformational changes in the protein and the DNA (or RNA) substrate. The color codes of the protein are the same as in Fig. 1. At the initial step in the reaction (A), the protein is bound to the ssDNA tail but does not bind the duplex region of the DNA. Upon binding ATP (B), there is a conformational change in the protein, and the duplex region binds to domains 1B and 2B with a concomitant unwinding of several base pairs at the junction. Finally (C), following hydrolysis of ATP, the protein conformation returns to that in (A) as the protein translocates along the ssDNA tail by one base and releases the DNA duplex. (b) Cartoon demonstrating the alternation in affinity for ssDNA of domains 1A and 2A during translocation. An open hand represents a loose grip on the DNA, and a closed hand a tighter grip. (A–C) correspond to those in (a) (Adapted from Velankar et al., 1999).

Another well studied SF1 helicase is Rep from *Escherichia coli*. In 1997, two crystal structures of binary and ternary complexes of Rep helicase bound to ssDNA or ssDNA and ADP were determined to a resolution of 3.0 Å and 3.2 Å, respectively (Korolev et al., 1997). The overall structure of Rep is similar to PcrA, consisting of fours domains with two parallel “RecA-like” cores, 1a and 2a and two additional domains, 1b and 2b (Fig. 1. 21). Interestingly, the asymmetric unit in the crystal contains two Rep monomers differing from each other by a large reorientation of domain 2b, with a revolution of 130° about a hinge region, hence designate as “open” and “close” conformer (Korolev et al., 1997) (Fig. 1. 21).
Fig. 1.21 Structure of Rep helicase.

Open and close conformers are shown in (A) and (B) respectively. Four domains (1a, 1b, 2a, 2b) are colored in green, yellow, red and cyan, respectively. The bound ADP is colored in purple.

Such a large conformation change and a bound ADP found in the close conformer suggested that these maybe coupled to translocation of the Rep dimer along DNA (Korolev et al., 1997). These studies have led to the proposal that the Rep dimer unwinds DNA in an “active rolling” mechanism (Korolev et al., 1997). In the active rolling model, the dimeric helicase interact with both dsDNA and ssDNA. Each subunit alternates binding to dsDNA as the dimer translocates when one subunit releases ssDNA and rebinds to dsDNA. Translocation along ssDNA is coupled to ATP binding (Fig. 1.22) (Tuteja and Tuteja, 2004).
Fig. 1. Active rolling model for DNA unwinding and translocation by the Rep DNA helicase dimer.

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

1.10.2.2 Structures and putative mechanisms of SF2 helicases

According to the classification of helicases proposed by Gorbalenya and Koonin, 1993, helicases of Flaviviridae belong to SF2 superfamily. HCV NS3 helicase is the first solved structure from SF2 superfamily. In contrast to PcrA and Rep, HCV NS3 helicase only has three domains, with two parallel “RecA-like” cores and an additional domain that is composed mainly of α helices (Yao et al., 1997; Kim et al., 1998; Cho et al., 1998; Mackintosh et al., 2006) (Fig. 1. 25 B). The seven motifs of SF2 helicases (Gorbalenya and Koonin, 1993) are located in domains 1 and 2 and the interdomain clefts. The molecular surface is punctuated by a deep groove separating the second domain from the rest of the molecule (Yao et al., 1997).

The structure reported by Yao et al., 1997 contains two HCV NS3 helicase monomers in the asymmetric unit. Comparison of crystallographically-independent molecules shows that rotation of domain 2 involves conformational changes within a conserved TATPP sequence (motif III) and untwisting of an extended antiparallel β sheets. Location of the TATPP sequence at the end of an NTPase domain structurally homologous to the “switch region” of many NTP-dependent
enzymes offers the possibility that domain rotation is coupled to NTP hydrolysis in the helicase catalytic cycle (Yao et al., 1997).

The structure reported by Cho et al., 1998 contains only one HCV NS3 helicase monomer and crystallized in space group P3\(_1\)21. Intriguingly, a channel is formed between a pair of symmetry-related molecules by extensive crystal packing interactions (Cho et al., 1998). A stretch of single-stranded RNA can be modeled into the interdomain cleft formed between RNA binding domain (domain 2) and the rest and continuously through the channel (Cho et al., 1998). These observations suggest that the dimer is likely to be the functional form that unwinds double-stranded RNA processively by passing one strand of RNA through the channel and the other outside of the dimer. A “descending molecular see-saw” model is then proposed (Cho et al., 1998) (See figure legend of Fig.1. 23).

![Fig.1. 23 Illustration of “descending molecular see-saw” mechanism of processive duplex unwinding by HCV NS3 helicase.](image)

(A), a helicase molecule \(\alpha\) binds the single-stranded RNA. (B), a functional dimer is formed by the binding of another helicase molecule \(\beta\). The dimer is...
stabilized by the interaction of ssRNA with the RNA binding motifs of α and β. (C), NTP hydrolysis by α results in the detachment of the ssRNA and a rigid body rotation of the dimer along an axis at the RNA binding motif of α. As a result, the dimer translocates along the ssRNA in the 3’ to 5’ direction, and the interdomain cleft of β binds the other portion of the ssRNA. (D), the dimer reaches the junction of ssRNA and dsRNA by repeated cycles of translocation. (E), in the same manner, the dimer translocates along the same strand of RNA. Energy required for the disruption of base pairings can be supplied by favorable interactions between the interdomain cleft and the ssRNA. One strand passing through the channel at the dimer interface is separated from the other strand hanging out of the dimer. (Adapted from Cho et al., 1998)

The structure reported by Kim et al., 1998 revealed HCV NS3 in complex with a single-stranded DNA oligonucleotide (dU8). Interestingly, the ssDNA lies in a groove between the first two domains (domain 1 and 2) and domain 3 with the 5’ end of the ssDNA resides at the interface of domains 2 and 3 and its 3’ end at the interface of domains 1 and 3. This orientation of ssDNA is roughly perpendicular to that of the ssRNA proposed by Yao et al., 1997 and Cho et al., 1998. Interactions between the ssDNA and enzyme are mostly confined to the DNA backbone, as would be expected for a nonspecific protein-nucleic acid complex, and are concentrated at the two ends of the oligonucleotide (Kim et al., 1998). In particular, Trp-501 stacks with the base of dU8 and Val-432 interacts with the dU4 base (Fig.1. 25 B). These two sidechains act as a pair of “book-ends”, defining a central binding cavity occupied by five nucleotides (Kim et al., 1998). A possible helicase mechanism performed by a monomeric enzyme was proposed (Kim et al., 1998) (See figure legend of Fig.1. 24).
The binding of polynucleotide by NS3 helicase in the absence of ATP leaves a large cleft between domains 1 and 2. Binding of ATP occurs with the $\beta$-phosphate binding to residues in the Walker A motif and the $\gamma$ phosphate mediated by Mg$^{2+}$ binding to the conserved acidic residues in Walker B motif (DECH). This results in the closing of the interdomain cleft and the binding of conserved arginines in motif VI to the ATP phosphates. Val-432 and Trp-501 disrupt base stacking at either end of the single-stranded region. Closure of the interdomain cleft leads to translocation of the single strand in the 5' to 3' direction and forces several bases to slip past Trp-501. Hydrolysis of ATP facilitates opening of the cleft and release of ADP. The orientation of Trp501 favors movement of the polynucleotide in only one direction such that opening of the gap results in net movement of the helicase in a 3'→5' direction (Adapted from Kim et al., 1998).

The crystal structure reported recently by Mackintosh et al., 2006 contains three HCV NS3 helicase monomers per asymmetric unit in complex with a single-stranded oligonucleotide long enough to accommodate binding of two monomers. In general, both proteins bind to the ssDNA in the same manner as that reported by (Kim et al., 1998). Several amino acid residues at the interface of the two HCV NS3 molecules mediate a protein-protein interaction between domains 2
and 3 of adjacent molecules (Mackintosh et al., 2006). Mutants in domain 3 disrupting the putative interaction were examined in vivo and in vitro using HCV subgenomic replicon and recombinant NS3 helicase, respectively (Mackintosh et al., 2006). The result showed that mutations led to a dramatic reduction in replication capacity without impairing the overall biochemical activities (Mackintosh et al., 2006). Inconsistencies in biological results suggest that the known biochemical properties associated with the helicase activity of NS3 do not reveal all of the likely biological roles played by NS3 during HCV replication (Mackintosh et al., 2006).

Recently, single molecule studies showed that the cyclic movement of HCV NS3 helicase is coordinated by ATP in discrete steps of 11 ± 3 base pairs, and that actual unwinding occurs in rapid smaller substeps of 3.6 ± 1.3 base pairs, also triggered by ATP binding, suggesting that HCV NS3 might move like an inchworm (Dumont et al., 2006).

Another helicase structure of flavivirus, yellow fever virus (YFV), was solved in complex with and without ADP (Wu et al., 2005). Like HCV NS3 helicase, YFV NS3 helicase also has three domains, with two parallel “RecA-like” cores which are similar to those of HCV and one additional domain 3 composed mainly of α helices which is strikingly different from that of HCV NS3 helicase (Wu et al., 2005) (Fig. 1. 25 B). A remarkable feature of YFV NS3 helicase is a groove formed by residues from three domains which is suitable to accommodate a single-stranded RNA substrate. Based on the common binding orientation observed in PcrA (Velankar et al., 1999) and HCV-dU8 (Kim et al., 1998)
complexes, Wu et al., proposed that a single-stranded RNA can proceed through the major interdomain cleft of helicase from domain 2 side towards domain 1 side of the protein (Wu et al., 2005). Furthermore, in the structure of ADP complex, the β phosphate binds the Walker A motif via extensive hydrogen bonds, while the adenine and ribose groups protrude from the NTP binding site, consistent with the lack of nucleotide specificity for the NTPase activity of flavivirus NS3 helicase (Warrener et al., 1993; Wu et al., 2005). Comparison of structures of apo and ADP complex revealed no significant domain movement between these two structures (Wu et al., 2005).

Fig. 1. Illustration of conserved motifs and structure of HCV NS3-dU8 helicase complex and YFV NS3-ADP complex.

(A). Diagram of the seven motifs defined by (Gorbalenya and Koonin, 1993), motifs I to VI are colored in cyan, purple, blue, orange, pink, yellow and wheat, respectively. (B). Schematic representation of HCV and YFV NS3 helicase structure. The conserved motifs are color-coded as in (A). Domain 1, 2, 3 are labeled. The bound ssDNA to HCV and ADP to YFV are illustrated as magenta and light blue.
sticks, respectively. Residues Trp-501 and Val-432, are also depicted as green sticks.

1.10.2.3 Structures and mechanisms of SF3 helicases

Several helicase crystal structures from SF3 have been reported, including the helicase of simian virus 40 (SV40) (Li et al., 2003; Gai et al., 2004) and Rep40 from adeno-associated virus (AAV) type 2 (James et al., 2003; 2004).

Unlike helicases of SF1 and SF2, structures of SF3 helicases possess only one “RecA-like” core (Fig. 1. 18, SF3) preceded by a bundle of α helices. The three sequence motifs (motif A, B, C) that define the SF3 superfamily (Gorbalenya and Koonin, 1993) are contained within the middle three β strands of the central sheet and the regions just flanking them (Hickman and Dyda, 2005) (Fig. 1. 26 A). SF3 helicases often form oligomeric rings, usually hexamers, and it is believed that DNA threads through a channel in the center of the ring (Neuwald et al., 1999). Furthermore, electron microscopy studies also confirmed that SV40 helicase can form an hexamer (Hickman and Dyda, 2005). The SV40 helicase was crystallized as a hexamer, representing the native functional state (Li et al., 2003). The SV40 helicase hexamer has a two-tiered domain arrangement, like a wedding cake, whereby the N-terminal α helices domains form a small ring on top of a larger ring of “RecA-like” cores (Fig. 1. 26 B). Interactions within an hexamer mainly comes from interfaces between neighboring N-terminal α helices and one side of the “RecA-like” core domain. The central channel formed by the N-terminal domains is wider than that formed by the “RecA-like” domains, suggesting that dsDNA can
be accommodated by the former but not the latter. Interestingly, hexamerization
place the Walker A motif of one monomer close to a conserved arginine on an
adjacent monomer, generating the expected arginine finger which is essential for
ATPase activity (Li et al., 2003; Abbate et al., 2004). In addition, a β hairpin that
precedes motif C points into the central channel, suggesting that DNA binding is
likely to be modulated by basic residues located at the finger tip (Fig. 1. 26 A).

Fig. 1. 26 Structure of SF3 helicases and SV40 hexamer.

(A) SV40 and Rep40 monomers. Motifs A, B, C are colored in red, light blue and
green, respectively. The β hairpin is colored in dark blue. A bound ADP in
Rep40 is illustrated as ball-and-stick. (B) SV40 hexamer, looking down through the
central channel. The N-terminal α helical domains (in gold) form a ring resting upon
a larger ring formed by “RecA-like” domains (in green) (Adapted from Hickman
and Dyda, 2005).

In the following years, the SV40 helicase was structurally characterized in the
presence of ATP and ADP (Gai et al., 2004), revealing significant conformational
differences in the hexameric assembly compared to the original apo structure. The
conformational changes predominantly reflect changes in interactions across
monomer-monomer interfaces. The six nucleotide-binding sites are either fully
occupied or all empty, and the presence of the nucleotide affects the size and shape
of the central channel, and the conformations of the six β hairpins that line the
channel (Gai et al., 2004; Hickman and Dyda, 2005) (Fig. 1.27).

Fig. 1.27 The change in channel openings of SV40 helicase hexamer.

To provide a clearer view of the nucleotide binding cleft at the hexamerization interface, only the "RecA-like" cores of the hexamer are shown. Each of the six monomers is in a different color. (A) The ATP bound hexamer structure. The six ATPs at the cleft between two monomers are in pink. (B) The ADP bound hexamer structure, showing ADP (pink) in the middle. (C) The apo hexamer structure (Adapted from Gai et al., 2004).

The structural differences of SV40 helicase in various nucleotide states detail the molecular mechanisms of conformational changes triggered by ATP binding and hydrolysis and reveal a potential mechanism of concerted nucleotide binding and hydrolysis. During these conformational changes, the angles and orientations between domains of a monomer alter, creating an "iris"-like motion in the hexamer. Additionally, six unique β hairpins on the channel surface move longitudinally along the central channel, possibly serving as a motor for pulling DNA into the hexamer for unwinding (Gai et al., 2004) (See figure legend of Fig. 1.28).
Fig 1. 28 Looping model of SV40 double hexamer unwinding.

The β hairpins move along the central channel in response to ATP binding (step 1), hydrolysis (step 2), and ADP release (step 3). (A) SV40 double hexamer with two ssDNA loops coming out from the side channels. Each hexamer contains a helicase domain (represented by two squares in pink) and an origin binding domain (ODB) (oval in light blue). The β hairpin structure is represented by two bars (in blue) within the helicase domain. The colored dots on the DNA (red, orange, black, and blue) are position markers for translocation. (B) A SV40 helicase hexamer corresponding to the left half of the double hexamer in (A) in the nucleotide-free state. (C) The movement of β hairpins upon ATP binding, which serves to pull dsDNA into the helicase for unwinding. The unwound ssDNA extrudes from the side channels. For clarity, only one hexamer is shown. (D) The β hairpins move back about halfway toward the nucleotide-free position after ATP hydrolysis. (E) The ADP is released from the SV40 hexamer, and the β hairpins return to the original nucleotide-free position (Adapted from Gai et al., 2004).
1.10.2.4 Structures and mechanisms of F4 helicases

The helicase domain of the replicative primase-helicase protein from bacteriophage T7 is the first helicase structure for the F4 family. It was crystallized in space group P6\(_1\) as a helical filament that resembles the *Escherichia coli* RecA protein (Sawaya et al., 1999). Like helicases of SF3, T7 helicase contains only one “RecA-like” core (Fig. 1.18 F4; Fig. 1.29 A) with four of the five conserved motifs (H1, H1a, H2, H3), which are implicated in nucleotide binding and hydrolysis, located at the C-terminal end of adjacent \(\beta\) strands (Sawaya et al., 1999; Singleton et al., 2000). When viewed in projection along the helical axis of the crystal, six monomers of T7 helicase resemble the hexameric rings seen in electronic microscopic images of the intact T7 helicase-primase (Egelman, 1996). dNTPs are found to bind the Walker A motif (H1 motif) of one monomer close to a conserved arginine of an adjacent monomer, generating the expected arginine fingers which are essential for NTPase activity (Sawaya et al., 1999). In addition, binding of nucleotides stabilizes the folded conformation of a DNA binding motif (H4) that is located near the central hole of the ring (Sawaya et al., 1999).
Fig. 1. 29 Structures of the T7 helicase.

(A) Overall fold of a T7 helicase monomer. The chain is colored from blue at the N-terminal to red at the C-terminal. A bound non-hydrolysable nucleotide analog is depicted as ball-and-stick. (B) Schematic representation of T7 helicase hexamer. The hexamer is stabilized via interactions between an N-terminal α helix and a pocket on the adjacent subunit. The four bound ADPNP molecules are depicted as ball-and-stick. (C) Ribbon diagram of T7 primase-helicase heptamer shows that the helicase domain (red) and associated DNA binding loops (yellow) are arranged in a ring with approximate 7-fold rotation symmetry. Primase domains are colored in grey. (Adapted from Singleton et al., 2000; Toth et al., 2003).

The structure of T7 helicase reported by Singleton et al., 2000 contained two independent half-hexamers in the asymmetric unit. The subunit interface is stabilized by interactions between an α helix in the N-terminal region and an adjacent subunit (Singleton et al., 2000) (Fig. 1. 29 B). Deviations from expected six-fold symmetry indicated that the structure is an intermediate on the catalytic pathway (Singleton et al., 2000). Interestingly, nonhydrolyzable ATP analogs were found to bind only four of the six possible nucleotide binding sites (Fig. 1. 29 B). Taken together, the structure suggested a “binding change” mechanism that explain how cooperative binding and hydrolysis of nucleotides are coupled to conformational changes in the ring that most likely accompany duplex unwinding (Singleton et al., 2000) (See figure legend of Fig. 1. 30).
In step I, Blue (B) subunits contain bound NTP, yellow (A) have NDP + Pi, and magenta (C) contain no bound nucleotide (empty). To progress to step II, hydrolysis of the bound NTPs results in conformational changes around the ring such that the bound NDP + Pi dissociates from the protein and the empty sites are now able to bind NTP. The same events are repeated in order to progress to step III, but take place at different subunits. Three cycles of NTP hydrolysis are shown to illustrate the nucleotide binding state of each subunit during successive cycles of the reaction and that all six NTP sites are utilized at different points during sequential cycles of hydrolysis. Therefore, ATP hydrolysis can be regarded as a ripple going around the ring without requiring a rotation of the ring itself (Adapted from Singleton et al., 2000).

Surprisingly, in 2003, Toth et al., reported that the structure of T7 primase-helicase was a heptamer. This heptamer has a crown-like shape, comprising of intimate packed helicase domains that are tiered with loosely arrayed primase domains (Toth et al., 2003) (Fig.1. 29 C). The advantage of heptameric isoform to accommodate double-stranded DNA in its central channel nicely explains its DNA remodeling activity (Kaplan and O’Donnell, 2002).
double-jointed structure of the primase-helicase permits a free range of motion for the primase and helicase domains that suggests how the continuous unwinding of DNA at replication fork can be periodically coupled to Okazaki fragment synthesis (Toth et al., 2003).

1.11 Aims of the project

Dengue fever is an important emerging public health concern, with several million viral infections occurring annually, for which no effective therapy currently exists. The NS3 helicase from dengue virus is a multifunctional protein endowed with helicase, nucleoside 5’-triphosphatase and 5’-terminal RNA triphosphatase activities. Thus, NS3 helicase plays an important role in viral replication and represents a very attractive target for the development of specific antiviral inhibitors. A better understanding of the 3D structure of dengue virus NS3 helicase domain should inform biochemical experiments to dissect the enzymatic mechanism and facilitate the design and development of specific antiviral drugs.
CHAPTER 2

MATERIALS AND METHODS

2.1 Overview

Dengue NS3 protein is a large multifunctional protein of 618 amino acids, endowed with protease, helicase, NTPase, as well as 5’-terminal RNA triphosphatase activities and plays an important role in viral polyprotein processing and genome replication (Lindenbach and Rice, 2001). The N-terminal 180 amino acids of NS3 comprises a serine protease domain with the protein NS2B acting as a membrane anchoring cofactor, necessary for proteolytic activity (Wengler et al., 1991; Chambers et al., 1993; Li et al., 1999; Yusof et al., 2000); its C-terminal domain, comprising amino acid 181 to 618, contains all the seven conserved motifs that is found in helicase SF1 and SF2 and is involved in viral RNA replication (Bartelma et al., 2002; Benarroch et al., 2004). This chapter describes the cloning, expression, purification of Den NS3 171-618 (DNS3H) and general phasing techniques used to solve the structure.

2.2 Cloning of DNS3H

2.2.1 Cloning strategy

pET32b (Novagen) was chosen for cloning because it contains a 11 kDa thioredoxin (Trx) tag which often increases both in the yield and solubility of the fusion protein (LaVallie et al., 1993). In this vector (Fig. 2.1 A), Trx was fused at the N-terminal end followed by a hexa-histidine tag, which facilitates purification by immobilized metal affinity chromatography (IMAC) and an enterokinase
cleavage site, allowing the cleavage of the recombinant protein. The gene coding for DNS3H was inserted between NcoI and HindIII. The selection of NcoI introduced three extra amino acids (Ala-Met-Ala) at the N-terminus of the recombinant protein after enterokinase cleavage, which turns to be important for crystallization (Fig. 2.1 B).

![Diagram of cloning strategies]

Fig. 2.1 Cloning strategies.

(A) The multiple cloning sites of pET32b (Novagen). (B) Illustration of the construct used to express the ns3 dengue helicase catalytic domain. The region upstream the NcoI site contains the gene encoding the Trx protein and a linker consisting of six histidine residues and an enterokinase cleavage site. The region downstream the NcoI site codes for the dengue ns3 helicase catalytic domain (residues 171-618 of the mature NS3 protein).
2.2.2 PCR, restriction enzyme digestion and ligation

The DNS3H gene, (Accession No AY037116; nucleotides 5032 to 6301), generously supplied by the Novartis Institute for Tropical Diseases (NITD), was amplified by Polymerase Chain Reaction (PCR) using the forward primer (5’-TAATTCCATGGCAAGCATCGAAGACAATCC-3’), and the reverse primer (5’-TAATAAGCTTTACTTTCTTCCAGCTGCAAA-3’). The underlined regions correspond to NcoI and Hind III (New England Biolab) sites respectively. The fragment was successfully amplified by using the protocol outlined below (Table 2.1; Fig. 2.2).

<table>
<thead>
<tr>
<th>Step</th>
<th>Tm (°C)</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Denaturation</td>
<td>95</td>
<td>3mins</td>
</tr>
<tr>
<td>2</td>
<td>95</td>
<td>45sec</td>
</tr>
<tr>
<td>3 Anneling</td>
<td>55</td>
<td>45sec</td>
</tr>
<tr>
<td>4 Extension</td>
<td>72</td>
<td>2.5mins</td>
</tr>
<tr>
<td>5 Go to 2 for 25 cycles</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6 Extension</td>
<td>72</td>
<td>10mins</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>+∞</td>
</tr>
</tbody>
</table>

*Fig. 2.2 PCR product of DNS3H (1.3Kb).*

The DNS3H PCR product and pET32b vector were double-digested by NcoI and HindIII for 2 hours at 37°C in buffer 2 (New England Biolab). After digestion, the PCR product and the vector were ligated using T4 ligase (New England Biolab) overnight, at room temperature. The ligation reaction was performed at a (vector:
insert) ratio of 1:3. The ligation mixture were then transformed into DH5α competent cells and spread onto Ampicillin LB (Luira Broth) agar plate. The plate was incubated at 37°C overnight.

2.2.3 Construct confirmation

Five single colonies were picked, used to inoculate a volume of 2 ml of Ampicillin LB medium, which was shaken overnight at 37°C. QIAprep Spin Miniprep Kit (Qiagen) was used to isolate the plasmid from bacterial colonies and restriction enzyme digestion (NcoI and HindIII) was used to confirm the presence of an insert in the construct (Fig. 2.3). The sequence of the insert was further confirmed using automatic DNA sequencing.

![Fig. 2.3 Confirmation of DNS3H with NcoI and HindIII double digestion. The 5.8 Kb and 1.3 Kb bands correspond to the pET32b vector and DNS3H respectively.]

2.3 Expression of Trx-DNS3H

The plasmid containing DNS3H was transformed into BL21-CodonPlus (Stratagene) E. coli competent cells for expression. A seed culture was prepared by inoculating one colony into 25 ml of LB with Ampicillin (100 μg ml⁻¹),
Chloramphenicol (25 μg ml⁻¹), incubated at 37 °C for overnight. 20 ml of seed culture was transferred to 1 litre LB with Ampicillin (100 μg ml⁻¹), Chloramphenicol (25 μg ml⁻¹) and incubated at 37 °C with shaking (220 rpm) until OD₆₀₀ reached 0.6-0.8. Isopropyl-β-D-1-Thiogalactoside (IPTG) was added to the final concentration of 0.4 mM. After overnight induction at 16 °C, cells were harvested by centrifugation at 8,000 g for 10 minutes at 4 °C. The culture supernatant was discarded and pellets were stored at -80 °C until use.

*Fig.2. 4* illustrates the resulting overexpression of Trx-DNS3H on a 10% SDS-PAGE gel.

![Fig.2. 4 Overexpression of Trx-DNS3H.](image)

Lane 1: Control; Lane 2: Induction, a band around 69 kDa in agreement with the calculated molecular weight of Trx-DNS3H fusion protein is present; Lane 3: Supernatant fraction; Lane 4: Pellet fraction.

2.4 Purification of DNS3H

2.4.1 Overview

Trx-DNS3H expressed from the pET-32b was subjected to several purification steps to generate protein suitable for crystallization purposes. These steps include
two IMAC, Anion exchange, Enterokinase cleavage and Gel filtration. Fig. 2. 5 summarizes the protocol used for purification.

![Flowchart of DNS3H purification protocol.](image)

**Fig. 2. 5 Flowchart of DNS3H purification protocol.**

### 2.4.2 Immobilized metal affinity chromatography (IMAC)

Cell pellet from 1 litre culture was resuspended in 40 ml lysis buffer (20 mM Tris-HCl, pH7.5, 0.3 M NaCl, 5% glycerol) and lysed by sonication (SONICS) for 20 minutes. The cell lysate was clarified by centrifugation at 30,000 g for 30 minutes at 4 °C and the supernatant was loaded onto a HisTrap HP column (Amersham Bioscience) pre-equilibrated with buffer A (see appendix). The column was thoroughly washed with buffer A and proteins were eluted using a linear imidazole concentration gradient from 0 to 100% of buffer B (see appendix). Recombinant proteins were eluted from a 50%-80% gradient corresponding to an imidazole concentration of 270-420 mM. The yield of recombinant protein was approximately 80 mg per litre culture. The elution profile and purity of Trx-DNS3H are shown on Fig. 2. 6.
2.4.3 Anion exchange chromatography

The eluate from IMAC was loaded onto a HiPrep 16/10 Q Sepharose Fast Flow (Amersham Bioscience) pre-equilibrated with buffer C (see appendix). The column was thoroughly washed with buffer C and proteins were eluted using a linear NaCl gradient.
concentration gradient from 0 to 100% of buffer D (see appendix). Recombinant protein was eluted from 50%-80% gradient. The elution profile and purity of Trx-DNS3H are shown on Fig. 2. 7.

![Anion exchange chromatography of Trx-DNS3H](image)

**Fig. 2. 7** Anion exchange chromatography of Trx-DNS3H.

(A) Anion exchange elution profile of Trx-DNS3H; (B) Lane 1: Eluate from Anion exchange.

2.4.4 Enterokinase cleavage of Trx-DNS3H

The eluate was subsequently pooled and concentrated to 6 mg ml\(^{-1}\) using ultracentrifugation at 3,000 g (Centricon 10 kDa MWCO, Vivascience) and the buffer changed to buffer E (see appendix) for enterokinase cleavage. In order to
maximize the yield of DNS3H protein and avoid the undesirable secondary cleavage in the portion of DNS3H (Fig. 2.8.8B), different parameters of the reaction, including the time, the molar ratio of enterokinase to Trx-DNS3H, were optimized (Fig. 2.8A). The best conditions found for the enzymatic cleavage are: 1 ng of enterokinase to 0.5 mg Trx-DNS3H (corresponding to an enzyme: protein substrate molar ratio of 1:200) at 4 °C for 12 hours. Protease inhibitors cocktail (Sigma) were then added to the cleavage mixture to stop the reaction.

**Fig.2.8 Optimization of enterokinase cleavage.**

(A) Lane 1: Initial sample of Trx-DNS3H before enterokinase cleavage; Lane 2-3: 11 and 12 hours after incubation with enterokinase using an enzyme to substrate molar ratio of 1:200 at 4 °C; Lane 4-5: 11 and 12 hours after incubation with enterokinase using an enzyme to substrate molar ratio of 1:400 at 4 °C; Lane 6-7: 11 and 12 hours after incubation with enterokinase using an enzyme to substrate molar ratio of 1:800 at 4 °C. The molecular weight of Trx-DNS3H, DNS3H and thioredoxin tag, as indicated in the figure, are 69 kDa, 51 kDa and 18 kDa, respectively. (B) Western blotting against an anti-DNS3H polyclonal serum (kindly provided by NITD). Lane 1: Before enterokinase cleavage; Lane 2-3: 12
and 20 hours after incubation with enterokinase using an enzyme to substrate molar ratio of 1:200 at 4 °C. Red arrow indicates an over-digested byproduct.

2.4.5 Removal of the thioredoxin protein tag

The cleavage mixture was loaded onto a thin Econo-column (BioRad) containing 2 ml of Ni-NTA resin pre-equilibrated with buffer F (see appendix). The flow through and wash fractions containing DNS3H were collected and concentrated (Fig. 2.9).

![Fig. 2.9 Removal of thioredoxin.](image)

Lane 1: cleavage mixture; Lane 2: flow through and washing fractions; Lane 3: elution fraction containing thioredoxin tag.

2.4.6 Gel filtration chromatography

Proteins were further purified using a HiPrep Sephacryl 16/60 S-100 HR column (Amersham Bioscience) pre-equilibrated with buffer G (see appendix). DNS3H appeared at 48 ml, corresponding to a monomer as determined using a gel filtration low molecular weight (LMW) calibration kit (Amersham Bioscience). Fractions containing pure DNS3H were pooled and concentrated to 20 mg ml⁻¹ in buffer H (see appendix) and stored at -80 °C until use (Fig. 2.10).
2.5 Expression of Selenomethionylated DNS3H

A single colony was picked and inoculated to a volume of 2 ml LB and cultured at 37 °C, 250 rpm for 12 hours. A volume of 150 μl of overnight culture was inoculated to a 150 ml LB supplemented with Ampicillin (100 μg ml⁻¹),
Chloramphenicol (25 μg ml⁻¹) at 37 °C overnight. One litre of M9 media (see appendix) containing 0.2% glucose, 2 mM MgSO₄, 0.1 mM NaCl, 0.5×10⁻⁴ % (w/v) thiamine, and all 20 amino acids except Glycine, Alanine, Proline, Aspartic acid, Cystein and Methionine at concentration of 0.04 mg ml⁻¹ each was prepared. The next day, 50 ml of the overnight culture was added to 1 litre of M9 supplemented with Ampicillin (100 μg ml⁻¹), Chloramphenicol (25 μg ml⁻¹) and the culture grown at 37 °C until an OD₆₀₀ of 0.5~0.6. High concentrations of isoleucine, lysine, and threonine are known to block methionine biosynthesis in *E. coli* by inhibiting aspartokinases. Moreover, phenylalanine and leucine act in synergy with lysine (Doublié, 1997). Therefore, threonine, lysine hydrochloride, phenylalanine at concentration of 100 mg L⁻¹; leucine, isoleucine, valine at concentration of 50 mg L⁻¹; L-selenomethionine at 60 mg L⁻¹ were added to the M9 media and under continuous shaking for another 15 minutes. IPTG was added to a final concentration of 0.4 mM. After overnight culture at 16 °C, cells were harvested by centrifugation at 8,000 g for 10 minutes at 4 °C. Supernatant was then discarded and pellets were stored at -80 °C until use (*Fig. 2. 11*).

![Fig. 2. 11 Overexpression of Selenomethionyl Trx-DNS3H.](image)

*Lane 1: Control; Lane 2: Induction. Black arrow indicates the fusion protein.*
2.6 Purification of Selenomethionylated DNS3H

The purification protocol for the Selenomethionyl DNS3H (SeMet DNS3H) was the same as for the native DNS3H except for the addition of 10 mM 2-mercaptoethanol to avoid oxidation of selenomethionine, which affects the magnitude and energy of the anomalous signal (Sharff et al., 2000). The final purified SeMet DNS3H is shown in Fig. 2.12.

![Fig. 2.12 Purification of SeMet DNS3H. Lane 1: Final purified SeMet DNS3H](image)

2.7 Mass spectrometry of native and SeMet DNS3H

Mass spectrometry was performed using a LCQ-Deca XP Plus ion trap mass spectrometer (Thermo Finnigan, San Jose, CA) to verify the incorporation of Selenomethionine. The mass of native DNS3H is 51,740 Da which differs from the theoretical mass by 0.37%. The mass of SeMet DNS3H is 52309 Da. The difference between native and SeMet is 569 and corresponds to 16.7 selenomethionine matching almost exactly the number of 16 methionine present in
the native DNS3H protein (Fig. 2.13).

![Fig. 2.13 Mass spectrometry of native and SeMet DNS3H.](image)

The blue arrow indicates proteins with incomplete incorporation of selenomethionine.

### 2.8 Phasing (Single-wavelength Anomalous Dispersion)

#### 2.8.1 Phase ambiguity intrinsic to SAD

When anomalous scattering is present, the structure factor of a reflection \((hkl)\) is expressed as

\[
F(h) = \sum_{j=1}^{N} (f_j^o + f_j^\lambda + if_j^\alpha) \exp(i2\pi h r_j) \tag{1}
\]

where \(f_j^o\) is the invariant normal atomic scattering factor, \(f_j^\lambda\) is the real part and \(f_j^\alpha\) is the imaginary part of the atomic scattering factors as a function of wavelength. (1) can be rewritten as

\[
F^+ = F^o + F^r + F^\prime \tag{2}
\]
and its complex conjugate of $F^-$ as

$$F^- = F(-h) = F^o + F^r + F^-,$$

(3)

from (1) (2) and (3), we have

$$F^*(-(-h)) \equiv F^{*-} = F^o + F^r - F^-,$$

(4)

where $F^{*-}$ is the conjugate complex of $F^-$. The relationship of (2), (3) and (4) is illustrated in Fig. 2.14.

Subtracting (4) from (2) we obtain: $F^+ - F^{*-} = 2F^-.$

(5)

Equation (5) defines a phase triangle. Because we can have an estimate of the amplitude of Friedel pairs ($F^+$ and $F^-$) by $|F^+(h)|^2 = I^+(h)$ and $|F^-(h)|^2 = I^-(h)$, respectively, from which the heavy atom positions can be deduced via an anomalous difference Patterson function, $F^-$ can also be calculated using the expression $F^-(h) = i \sum_{j=1}^{N} f_j \exp(2\pi i h r_j)$.

(6)

The phase triangle can be illustrated as in Fig. 2.15, for a given reflection, this leads to two possible phases, one correct and the other incorrect. The averaged amplitude given by (7) (Fig. 2.15) is normally used for calculating electron-density maps.
Fig. 2. 15 The phase doublet of SAD.

The green circle is centred at the end of the vector $F^*$ with radius $|F^-|$ or $|F^*|$. The red circle is centred at the end of the vector $-F^*$ with radius $|F^*|$. $F^*$ is the structure factor contributed from the imaginary part of the heavy atom. $\varphi^*$ is the phase of $F^*$. $F^*$ and $F^-$ are structure factors of a Friedel pair $(h, -h)$. The two possible phases triangles lead to two possible phases, $\varphi^* = \varphi^* + |\Delta \varphi|$ or $\varphi^- = \varphi^- - |\Delta \varphi|$ for the averaged structure factor $< F >$.

The averaged structure factor is defined by

$$< |F| \geq (|F^*| + |F^-|)/2$$

The two possible phases associated with $< F >$ can be expressed as

$$\varphi = \varphi^* \pm |\Delta \varphi|$$

where $\varphi^*$ is the phase of $F^* (h)$ which can be calculated from the heavy atom position and $|\Delta \varphi|$ is the absolute difference between $\varphi^*$ and the phase associated with $< F >$.

Hence, the phase ambiguity in SAD can also be considered as the sign ambiguity of $\Delta \varphi$ (Wang et al., 2004). The problem then arises as which phase to choose, this requires a consideration of phase probabilities.
2.8.2 Bimodal phase probability distribution from SAD

In reality, there are errors associated with the measurements of the structure factors and in the heavy atom positions and their occupancies, so the vector triangle seldom closes. David Blow and Francis Crick introduced a concept of “lack-of-closure” (ε) and used it to define a phase probability distribution (Blow and Crick, 1959). By assuming that all the errors reside in $F_{PH}(calc)$, the errors follow a Gaussian distribution (Fig. 2.16).

Given the magnitude $<|F|>$ (7) and the structure factor contributed by the imaginary part of the heavy-atom (6), the relative probability of a possible phase of $<F>$ can be estimated from the agreement between the observed anomalous difference $(\Delta_{ANO}^0)$ and the anomalous difference calculated using the phase $[\Delta_{ANO}^C(\varphi_h)]$.

$$P_{SAD}(\varphi_h) \propto \exp\left\{-\left[\Delta_{ANO}^0 - \Delta_{ANO}^C(\varphi_h)\right]^2 / (E_{ANO}^2 + 2\sigma_{ANO}^2)\right\},$$

where $\sigma_{ANO}^2$ is the variance in the measurement of $\Delta_{ANO}^0$ and $E_{ANO}^2$ the square of the lack-of-closure error.

![Fig. 2.16 Schematic plot of phase distribution curve from phasing of SAD data.](image)

(A) The $\varphi^*$ is the best phase and $M$ the centroid of this distribution, denoted as $F_{best}$, whose amplitude is the native amplitude $|F_p|$ weighted by the figure of merit,
2.8.3 Phase improvement

Due to the inherent ambiguity and experimental errors, experimental phases from SAD are usually insufficiently accurate to directly yield a completely interpretable electron density map. In fact, experimental phases are often only the starting point for phase improvement using a variety of methods of real space density modification, which are based on some prior knowledge of structure. Solvent flattening, histogram matching and non-crystallographic averaging are the main techniques used to modify electron density and improve phases. Solvent flattening removes negative electron density and sets the value of electron density in protein regions to a typical value of 0.43 e Å⁻³, in contrast to solvent region of 0.33 e Å⁻³. Histogram matching alters the values of electron density points to match an expected distribution of electron density values. Non-crystallographic symmetry averaging is the most powerful technique which imposes equivalence on electron density values when there is more than one copy of a molecule present in the asymmetric unit or in different crystals (Taylor, 2003). Other techniques such as multi-resolution modification, skeletonisation (Baker et al., 1994) and Sayre’s equation (Sayre, 1974) can be also useful.
CHAPTER 3

RESULTS AND DISCUSSION

3.1 Crystallization of DNS3H

Native DNS3H was screened for crystallization conditions using the Hampton screen (Hampton Research) kit 1, 2 and PEG/ION at concentration of 20 mg ml\(^{-1}\). 1 \(\mu\)l of protein solution was mixed with 1 \(\mu\)l of the reservoir solution using the hanging-drop vapour diffusion method. After 1 day incubation at 18°C, clusters of needle-like crystals were obtained in the condition containing 0.1 M sodium cacodylate (pH6.5), 0.2 M (NH\(_4\))\(_2\)SO\(_4\), 30% polyethylene glycol 8000. To optimize the crystallization condition, the pH, protein concentration and polyethylene glycol 8000 concentration were screened. For both native and SeMet DNS3H, the best condition found were 0.1 M MES (morpholine- ethanesulfonic acid, pH 6.5), 0.2 M (NH\(_4\))\(_2\)SO\(_4\), 14% polyethylene glycol 8000 (1mM Manganese Chloride for Mn\(^{2+}\) complex) with a protein concentration of 5mg ml\(^{-1}\) at 18°C. After streak seeding, a technique where small nuclei or micro-crystals are transferred with the help of a whisker to a protein solution that is optimized for crystal growth, crystals of the native (as well as the SeMet) protein grew as thin elongated plates over 5-7 days to dimensions of approximately 0.02 x 0.3 x 0.1 mm\(^3\) (Fig.3. 1).

![Fig.3. 1 DNS3H crystals.](image)

Native DNS3H crystals after streak seeding.
3.2 Data collection and reduction

For data collection, the native, SeMet or Mn$^{2+}$ complex crystals were soaked in a cryoprotecting solution of 0.1 M MES, pH 6.5, 0.2 M (NH$_4$)$_2$SO$_4$, 14% Polyethylene glycol 8000, 25% glycerol, (1 mM Manganese Chloride for Mn$^{2+}$ complex) before being mounted and cooled to 100K in a nitrogen gas stream (Oxford cryosystems). Diffraction intensities were recorded on tunable beamline ID14-4, which is suitable for Multiple-wavelength Anomalous Dispersion (MAD) data collection, at the European Synchrotron Radiation Facility (Grenoble, France) on an ADSC charge-coupled device detector. Even though an attenuated beam of dimensions 0.125 by 0.050 mm$^2$ was used in order to minimize radiation damage, the data collected at remote and edge wavelength suffered dramatically from radiation damage and were therefore not included in phase determination. Data integration, scaling, and merging of the intensities were done by programs from the CCP4 suite (CCP4, 1994). The native, SeMet and Mn$^{2+}$ complex crystals diffracted to 2.4 Å, 2.8 Å and 2.75 Å, respectively. Data collection statistics are summarized in Table 3.1.
<table>
<thead>
<tr>
<th></th>
<th>Native</th>
<th>SeMet peak</th>
<th>Mn$^{2+}$ Complex</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wavelength (Å)</strong></td>
<td>0.97600</td>
<td>0.97938</td>
<td>0.97927</td>
</tr>
<tr>
<td></td>
<td>$a=54.9$</td>
<td>$a=54.7$</td>
<td>$a=54.26$</td>
</tr>
<tr>
<td><strong>Cell parameters (Å), P2$_1$</strong></td>
<td>$b=178.1$</td>
<td>$b=165.9$</td>
<td>$b=177.53$</td>
</tr>
<tr>
<td></td>
<td>$c=55.4$</td>
<td>$c=54.7$</td>
<td>$c=55.09$</td>
</tr>
<tr>
<td><strong>β (º)</strong></td>
<td>101.3</td>
<td>101.8</td>
<td>101.57</td>
</tr>
<tr>
<td><strong>Resolution range (Å)</strong></td>
<td>29-2.41</td>
<td>29-2.8</td>
<td>20-2.75</td>
</tr>
<tr>
<td><strong>No. of observed reflections</strong></td>
<td>265,358</td>
<td>174,490</td>
<td>125,866</td>
</tr>
<tr>
<td><strong>No. of unique reflections</strong></td>
<td>37,606 (16,898)</td>
<td>45,846 (8,517)</td>
<td>60,408 (9,784)</td>
</tr>
<tr>
<td><strong>Completeness</strong></td>
<td>92.5% (82.5%)</td>
<td>88.5% (60.4%)</td>
<td>96.6% (97.3%)</td>
</tr>
<tr>
<td><strong>Multiplicity</strong></td>
<td>7.0 (6.7)</td>
<td>3.8 (3.7)</td>
<td>2.08 (2.11)</td>
</tr>
<tr>
<td><strong>$R_{merge}$</strong></td>
<td>0.062 (0.214)</td>
<td>0.067 (0.218)</td>
<td>0.075 (0.357)</td>
</tr>
<tr>
<td><strong>I/σ(I)</strong></td>
<td>20.18 (8.82)</td>
<td>15.21 (6.41)</td>
<td>11.8 (3.23)</td>
</tr>
<tr>
<td><strong>f' / f''</strong></td>
<td>-8.1 / 5.7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^1 R_{merge} = \Sigma_h \Sigma_i |I_{hi} - \langle I_h \rangle| / \Sigma_h \Sigma_i I_{hi}$, where $I_{hi}$ is the $i$th observation of the reflection $h$, while $\langle I_h \rangle$ is its mean intensity.

$^2$ Values of $f'$ and $f''$ where estimated from a scan of the absorption edge using program CHOOCH (Evans and Pettifer, 2001).
3.3 Phasing of DNS3H

The SAD method was used to solve the DNS3H structure because of radiation damage. Out of the 32 selenium atoms present in the two molecules of the asymmetric unit, 28 could be located using program SOLVE at 3 Å resolution (Terwilliger, 2003). The statistics of SAD phasing are summarized in Table 3.2.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>MEAN</th>
<th>SD</th>
<th>VALUE</th>
<th>Z-SCORE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pattersons</td>
<td>3.84</td>
<td>0.500</td>
<td>8.08</td>
<td>8.48</td>
</tr>
<tr>
<td>Cross-validation Fourier</td>
<td>8.02</td>
<td>2.47</td>
<td>137</td>
<td>52.2</td>
</tr>
<tr>
<td>NativeFourier CC x100</td>
<td>7.86</td>
<td>0.744</td>
<td>23.2</td>
<td>20.7</td>
</tr>
<tr>
<td>Mean figure of merit x100</td>
<td>0.00</td>
<td>5.00</td>
<td>42.4</td>
<td>8.47</td>
</tr>
<tr>
<td>Correction for Z-scores</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-21.8</td>
</tr>
<tr>
<td>Overall Z-score value</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>68</td>
</tr>
</tbody>
</table>

An initial map was calculated and modified with program RESOLVE (Terwilliger, 2003), using the heavy atom positions to locate the noncrystallographic symmetry axis relating the two molecules present in the asymmetric unit. Program SHARP (delaFortelle and Bricogne, 1997) was used to locate other three selenium atoms and a new set of single anomalous dispersion phases were calculated to 2.8 Å (Table 3.3). One methionine introduced by NcolI at the N-terminal is disordered and invisible. The electron density was then modified by program SOLOMON/DM (CCP4, 1994). The resulting map was of extremely
good quality and allows the tracing of about 400 residues manually for one molecule in the asymmetric unit.

Table 3.3 Phasing statistics

<table>
<thead>
<tr>
<th></th>
<th>SeMet peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of sites</td>
<td>31</td>
</tr>
<tr>
<td>( R_{\text{cullis}} ) (acentrics) (^1)</td>
<td>0.60</td>
</tr>
<tr>
<td>Phasing power (^2)</td>
<td>1.915</td>
</tr>
<tr>
<td>Figure of merit (^3)</td>
<td>0.41/0.87</td>
</tr>
<tr>
<td>20-2.8 Å</td>
<td></td>
</tr>
</tbody>
</table>

\(^1\) \( R_{\text{cullis}} = \frac{\sum_h \left| F^+ - F^- \right| - \left| F_{H(\text{calc})} \right|}{\sum_h \left| F^+ - F^- \right|} \) where \( F_{H(\text{calc})} \) is the calculated heavy atom structure factor. The summation is over acentric reflections only.

\(^2\) Phasing power is the r.m.s heavy atom structure factor divided by the r.m.s. lack of closure.

\(^3\) Figure of merit are given before and after real space density modification, respectively.

3.4 Model building, refinements and quality of the model

Model building was carried out with the program O (Jones, 1991). SeMet residues were used as markers to assist amino acid sequence assignment. The SeMet derivative structure was partially refined with CNS (Brunger, 1998) using ncs and experimental single anomalous dispersion phases restraints. The best
defined regions of the two independent molecules in the asymmetric unit were initially constructed. Once a fragment was constructed in one molecule, the ncs operator was applied to generate its counterpart in the second molecule. The partial model was subsequently used as a model to solve the native structure at a resolution of 2.4 Å by molecular replacement using program MOLREP (CCP4, 1994). Iterative 2Fo-Fc and Fo-Fc maps were calculated to complete the model. Finally, molecular replacement was performed using the native model as a guide to complete the SeMet structure model. Water molecules were picked using the Fo-Fc map at a level of 3.0σ by CNS (Bruner, 1998). The geometry of the final model was checked with PROCHECK (Laskowski et al., 1993). The refinement statistics are summarized in Table 3.4.

The native and SeMet proteins both crystallized in space group P2_1, with a slight variation in one unit cell dimension (Table 3.1). All crystal forms contain two molecules in the asymmetric unit with similar crystal packing. Among all three structures, one of the two molecules in the asymmetric unit has a lower average temperature factor and is better ordered. Superposition of all six molecules reveals an average r.m.s deviation of 0.81Å. Differences are located at the N-terminal ends, within the phosphate binding loop and in the loops exposed to the solvent. Residues 244 to 251 within domain 1 in both native and SeMet structures are disordered in our model. In the Mn^{2+} complex, residues 244 to 255 in molecule B and residues 243 to 255; 386 to 398; 411 to 417 in molecule A are disordered in our structure. The equivalent residues of 244 to 254 in DNS3H form the solvent accessible strand β4 of the seven-stranded β-barrel in the hepatitis C virus NS3 structure (Kim et al., 1998) but in the structure of yellow fever NS3 helicase, this
stretch of sequence form a solvent-exposed loop and involved in crystal packing (Wu et al., 2005). A number of charged side chains on the surface of the protein are also not visible in the electron density map and have been omitted from the final model. Figures were produced using program Pymol (DeLano, 2002).
<table>
<thead>
<tr>
<th></th>
<th>Native</th>
<th>SeMet</th>
<th>Mn$^{2+}$ complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution range (Å)</td>
<td>20.0 - 2.4</td>
<td>20.0 - 2.8</td>
<td>20.0 - 2.75</td>
</tr>
<tr>
<td>Intensity cutoff (F/σ(F))</td>
<td>0.</td>
<td>0.</td>
<td>0.</td>
</tr>
<tr>
<td>No of reflections: completeness (%)</td>
<td>98.1</td>
<td>99.0</td>
<td>97.7</td>
</tr>
<tr>
<td>Used for refinement</td>
<td>35671</td>
<td>21987</td>
<td>24536</td>
</tr>
<tr>
<td>Used for Rfree calculation</td>
<td>1884</td>
<td>1198</td>
<td>1243</td>
</tr>
<tr>
<td>No of non hydrogen atoms</td>
<td>3473, 3516</td>
<td>3494, 3516</td>
<td>3279, 3489</td>
</tr>
<tr>
<td>Protein missing residues##</td>
<td>(10, 10)</td>
<td>(10, 10)</td>
<td>(32, 13)</td>
</tr>
<tr>
<td>SO$_4^{2-}$</td>
<td>0</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Mn$^{2+}$</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Water molecules</td>
<td>195</td>
<td>134</td>
<td>124</td>
</tr>
<tr>
<td>Rfactor§ (%)</td>
<td>22.9</td>
<td>21.4</td>
<td>21.4</td>
</tr>
<tr>
<td>Rfree# (%)</td>
<td>26.7</td>
<td>26.2</td>
<td>25.2</td>
</tr>
<tr>
<td>Rms deviations from ideality</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.0069</td>
<td>0.0081</td>
<td>0.0073</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.29</td>
<td>1.41</td>
<td>1.37</td>
</tr>
<tr>
<td>Ramanchandran plot</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>most favoured regions (%)</td>
<td>83.1</td>
<td>82.2</td>
<td>85.4</td>
</tr>
<tr>
<td>additional allowed regions (%)</td>
<td>16.9</td>
<td>16.8</td>
<td>14.0</td>
</tr>
<tr>
<td>Overall G factor*</td>
<td>0.21</td>
<td>0.18</td>
<td>0.21</td>
</tr>
<tr>
<td>PDB codes</td>
<td>2BMF</td>
<td>2BHR</td>
<td>-</td>
</tr>
</tbody>
</table>

## Values are given for molecule 1 and 2 respectively

§ $R_{factor} = \Sigma | |F_{obs}| - |F_{calc}| | / \Sigma |F_{obs}|.$

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

* G factor is the overall measure of structure quality from PROCHECK (Laskowski et al., 1993).
3.5 The structure of DNS3H

3.5.1 Overall structure

The structure of DNS3H reveals a three-lobed flattened structure (Fig. 3.2), comprising three roughly equal size domains with overall dimensions of about 60 Å by 60 Å by 35 Å. A significant structural feature is a long tunnel that runs across the center of one face of the protein (Fig. 3.2).

Fig. 3.2 Structure of DNS3H.
(A) Overall fold of the DNS3H. Secondary structure elements are labelled. The same color code (and the relative orientation between the three domains) is also used in the topology diagram. The sulfate ion bound in the ATPase catalytic site is represented as sticks. (B) Folding diagram of the DNS3H. The location of the conserved SF2 helicases motifs is indicated with the same color code. The missing segment connecting strands $\beta_{2A}$ and $\beta_3$ in domain I comprising residues 244-253 is drawn as a broken line.

An amino-acid sequence alignment of the NS3 helicases from several members of the *Flaviviridae* is shown in Fig.3. 3. In contrast to SF1 helicases, which comprise four structural domains, two being present as insertions within the tandem core structure (Subramanya *et al.*, 1996; Korolev *et al.*, 1997), the DNS3H reveals three structural domains of about 150 amino-acids each, separated by a series of clefts following each other in sequence, like in the HCV NS3 helicase structure (Cho *et al.*, 1998, Kim *et al.*, 1998). Domain I (residues 181-326) and domain II (residues 327-481) show little sequence identity with each other, but are structurally similar being composed of a large central six stranded parallel $\beta$-sheet with a pronounced twist, flanked by four $\alpha$-helices (*Fig.3. 2 B*). Domain III (residues 482-618) consists a bundle of four approximately parallel $\alpha$-helices ($\alpha_{1''}$, $\alpha_{3''}$, $\alpha_{4''}$, $\alpha_{7''}$), surrounded by three shorter helices ($\alpha_{2''}$, $\alpha_{5''}$, $\alpha_{6''}$), and augmented by a double-stranded antiparallel $\beta$ sheet that exposed to the solvent. The major contact between domains include a long $\beta$-hairpin ($\beta_{4A'}$, $\beta_{4B'}$) that extends largely from domain II into domain III and interactions between helix $\alpha_3$ of domain I with $\alpha_{1''}$ and $\alpha_{2''}$ of domain III (*Fig.3. 2 A*). A superposition of the six independent molecules in the three crystal reveals no hinge motion between
domains, suggesting a rather rigid structure in the absence of nucleic acid or ATP.

**Fig. 3.3** Structural alignments of Flaviviridae helicase domains.

Alignment was produced using the program T-Coffee (Notredame et al., 2000).
After cleavage with enterokinase, the mature protein DNS3H which was crystallized has three additional amino acids (Ala, Met, Ala) at its N-terminal end that come from cloning. Sequences of JEV (Japanese Encephalitis Virus, strain JaOArS982, NP_059434), WNV (West Nile Virus, strain B956, AAT02759), TBV (Tick-borne Encephalitis Virus, strain western subtype, NP_043135), YFV (Yellow Fever Virus, strain 17D, NP_041726), HCV (Hepatitis C virus, strain H77, NP_671491) were obtained from GenBank. Secondary structure elements of DNS3H are displayed above the sequence alignment. The conserved motifs among SF2 helicases are indicated with a purple (motif I), blue (motif II), orange (motif III) and yellow background (Ia, IV, V, VI) and labeled.

A comparison with the HCV NS3 structures is shown in Fig. 3. Domains I and II share respectively 19% and 20% sequence identity with their counterparts in HCV NS3 and have a similar structure. After superposition of the individual domains, the r.m.s. deviations are 2.3 Å and 1.8 Å for domain I (114 equivalent Cα atoms) and II (132 equivalent Cα) respectively. Domain III (residues 482-618) differs most between Den and HCV, consistent with the lack of detectable sequence identity between the C-terminal ends of these two proteins.
The two molecules were superimposed based on their structurally conserved domains I and II, their domain III (top of the figure) bear no significant structural similarity.

A comparison with the YFV NS3 structures is shown in Fig. 3. 5. DNS3H and YFV NS3 share 48% sequence identity and have a similar structure. After superposition, the r.m.s. deviation is 2 Å for 416 Cα atoms. The hinge between domains 1 and 2, formed by motif III, is about 4.2° more closed in the Den helicase than in the structure of the YFV helicase. Significant difference in domain II involves a short helix $\alpha_3'$ and a loop between $\alpha_3'$ and $\beta_4'$, corresponding to residues Thr-408 to Phe-417 (motif V) (Fig. 3. 5). In domain III, the difference is more pronounced, except for half of the N terminal part of domain III, which superimposes well between two structures, the rest in domain III of Den helicase is about 10° more closed than in the structure of the YFV helicase with respect to the loop between $\alpha_4''$ and $\alpha_5''$ (Fig. 3. 2 A). This dramatic change in conformation
suggests that during the enzymatic cycle, the helicase might adopt different conformations, such as open and close states.

Fig. 3. 5 Superposition of the Cα-carbon atoms of DNS3H (red) and YFV helicases (blue).

The two molecules were superimposed based on sequence alignment of three domains. A close-up view of loop region (Thr-408 to Phe-417 for DNS3H in red and Thr-413 to Leu-422 for YFV NS3 in blue) is shown on the right panel. The window panel is a view with the molecule rotated by 90° around a vertical axis.

A search through the PDB (Protein Data Bank), using the DALI server (Holm et al., 1993) did not reveal any significantly homologous structure, indicating that domain III of DNS3H has a unique fold. This further illustrates that during
Chapter 3

evolution, the common structural core formed by the “RecA-like” tandem structures has been augmented with additional domains present as C-terminal additions for the SF-2 superfamily. Domain III, however, influences both the NTPase and helicase activity as demonstrated by the mutation of a single Arg residue within helix $\alpha_2''$ (Arg-513-Ala) which produces a defective helicase (Benarroch et al., 2004). Recently, residues 303-618 of DNS3H were shown to bind to the RNA dependent RNA polymerase (RdRp) NS5 (Brooks et al., 2002). This interaction might involve the C-terminal domain III of DNS3H. It will be interesting to determine whether the structural differences observed in the domain III, between the HCV and DNS3H, are correlated with a diverging mode of interaction with their respective RdRp.

3.5.2 Non-crystallographic dimer

Inside the crystal, the N-terminus of one monomer adopts an extended conformation, forming intermolecular contacts with residues of a loop between helix $\alpha_5'''$ and strand $\beta_1''$ of domain III in the neighbouring molecule (Fig.3. 6 A, C). Owing to the presence of the protease domain at its N-terminus, a similar interaction could not be formed by the full length NS3 protein. The other main intermolecular interaction involves residues 396-401 from domain II of one monomer, with residues from domain III of the other monomer (Fig.3. 6 A, B; Table 3. 5).
Fig. 3. 6 Major crystal contacts.

(A) Ribbon representation of three monomers in contact. Monomers A and B are colored in green and blue, respectively. (B&C) Close-up views of residues in contact. Hydrogen bonds and salt bridge interactions are shown as yellow dash. Residues of monomer A and B are shown as green and blue sticks and labeled in roman and italic style, respectively. Panel B is a view with the molecule rotated by 180º around a horizontal axis.
Table 3. 5 Major crystal contacts

<table>
<thead>
<tr>
<th>ResID1</th>
<th>Atom1</th>
<th>ResID2</th>
<th>Atom2</th>
<th>Distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met169A</td>
<td>N</td>
<td>Leu573B</td>
<td>O</td>
<td>3.12</td>
</tr>
<tr>
<td>Ala170A</td>
<td>N</td>
<td>Leu573B</td>
<td>O</td>
<td>2.95</td>
</tr>
<tr>
<td>Ala170A</td>
<td>O</td>
<td>Leu573B</td>
<td>N</td>
<td>2.50</td>
</tr>
<tr>
<td>Ser171A</td>
<td>OG</td>
<td>Tyr601B</td>
<td>OH</td>
<td>2.88</td>
</tr>
<tr>
<td>Ile172A</td>
<td>N</td>
<td>Gln571B</td>
<td>O</td>
<td>2.82</td>
</tr>
<tr>
<td>Glu173A</td>
<td>N</td>
<td>Asn569B</td>
<td>O</td>
<td>3.28</td>
</tr>
<tr>
<td>Glu173A</td>
<td>OE2</td>
<td>Asn569B</td>
<td>OD1</td>
<td>3.19</td>
</tr>
<tr>
<td>Pro176A</td>
<td>O</td>
<td>Asn569B</td>
<td>OD1</td>
<td>3.26</td>
</tr>
<tr>
<td>Glu177A</td>
<td>OE1</td>
<td>Asn569B</td>
<td>OD1</td>
<td>3.20</td>
</tr>
<tr>
<td>Lys396B</td>
<td>NZ</td>
<td>Ala552A</td>
<td>O</td>
<td>3.24</td>
</tr>
<tr>
<td>Lys396B</td>
<td>NZ</td>
<td>Glu553A</td>
<td>OE2</td>
<td>2.65</td>
</tr>
<tr>
<td>Thr399B</td>
<td>O</td>
<td>Arg526A</td>
<td>N</td>
<td>2.70</td>
</tr>
<tr>
<td>Thr399B</td>
<td>OG1</td>
<td>Ala552A</td>
<td>O</td>
<td>2.83</td>
</tr>
<tr>
<td>Asp401B</td>
<td>N</td>
<td>Arg524A</td>
<td>O</td>
<td>2.94</td>
</tr>
</tbody>
</table>

Oligomerization has been proposed as a means to provide helicases with multiple nucleic acid binding sites, such as HCV NS3 helicase (Levin et al., 1999; Serebrov et al., 2004), which would facilitate translocation of the protein along the strands. However, the dimer observed in our crystal differs from the one reported (Cho, et al., 1998). Thus, it is unclear whether the dimer present in our crystal structure merely derives from packing constraints or from a specific interaction. However, the relatively small surface area (1,050 Å²) buried in this interaction
would favour the first possibility (Bahadur et al., 2004). This monomeric form is consistent with our observation using size exclusion chromatography in Den helicase, as well as the observation of YFV helicase in solution using dynamic light scattering (Wu et al., 2005).

3.5.3 The NTPase active site

We obtained the crystal structure of DNS3H with a bound Manganese and a sulfate ion from the crystallization buffer located in the NTPase binding pocket next to the N-terminal end of helix α1 (Fig. 3.7), making close contacts with Arg-463 of motif VI, and residues protruding from the P-loop (motif I) and motif II. A network of solvent molecules are also buried in this pocket. In general, the Mn$^{2+}$ ion binds to DEAH motif via coordination of hydrogen bonds to O$_{ε1}$ Glu-285 (Walker B motif), O$_{γ1}$ Thr-200 (Walker A motif), oxygen of SO$_4^{2−}$ and 4 water molecules. The sulfate ion binds to the Walker A motif via hydrogen bonds to the backbone NHs of Gly-196 to Thr-200 (GAGKT) and to the side chain of Lys-199. The overall orientations of the residues in this cavity match very well with equivalent residues of the YFV helicase bound to a ADP (Wu et al., 2005). As shown in Fig. 3.7, the model places the base and ribose moieties bulging out from the NTP binding cavity, consistent with the lack of nucleotide specificity for the NTPase activity of flavivirus NS3 helicases (Warrener et al., 1993). In addition, the site is also ideally suitable to accommodate the 5′-triphosphate end of an RNA substrate, consistent with the evidence that the NTPase active site also catalyzes the RNA triphosphatase reaction (Benaroch et al., 2004).
Fig. 3. 7 Superposition of the DNS3H $\text{Mn}^{2+}$-$\text{SO}_4^{2-}$ (green sticks) and YFV helicase ADP complex (yellow sticks).

The two molecules were superimposed based on Walker A and Walker B motif. Amino acid numbering with and without parenthesis corresponds to the numbering in dengue and yellow fever helicase, respectively. $\text{Mn}^{2+}$-$\text{SO}_4^{2-}$ ions are shown in Fourier difference omit map contoured at $5\sigma$.

Superposition of the DNS3H Mn$^{2+}$ and YFV helicase ADP complex reveals that the distance of the sulfate to $\beta$ phosphate of ADP is 2.7 Å, close to the bond length 2.8 Å between $\gamma$ and $\beta$ phosphate of ATP. Wu et al. stated that, in the YFV helicase and ADP complex, neither the $\gamma$-phosphate nor Mg$^{2+}$ was observed in the NTPase binding pocket due to the hydrolysis of Mg$^{2+}$-ATP (Wu et al., 2005). The structure we present here complements the missing part of their model and gives new insight into the mechanism of NTPase of flavivirus helicase.

Superposition of DNS3H with PcrA DNA helicase bound to a non-hydrolyzable ATP analogue (Velankar et al., 1999) also gives some information on the NTPase mechanism (Fig. 3. 8). Overall, residues Lys-37, Thr-38, Asp-223, Glu-224, Gln-254, Arg-610, and Arg-287 of PcrA (Velankar et al., 1999) superimpose on their counterparts in DNS3H: Lys-199, Thr-200 of the P-loop (motif I), Asp-284,
Glu-285 (motif II) and Gln-456, Arg610, and Arg-463 of motif VI, respectively. These residues are likely to play similar roles during NTP hydrolysis. Upon the binding of ATP, residues surrounding the NTPase binding pocket might undergo certain conformational changes into a close state, with the side chain of Lys-199 contacting the β-phosphate of the ATP, Arg-463 or Arg-460 grabbing the γ-phosphate during transition state stabilization, and the strictly conserved Asp-284 and Glu-285 residues coordinating the divalent cation (Mg²⁺ or Mn²⁺). After ATP hydrolysis, ADP and γ-phosphate are released and helicase returns to its initial state (open state) (Velankar et al., 1999).

Fig. 3. 8 DNS3H active site.

View of the ATP binding pocket of DNS3H (yellow sticks) with the bound
Manganese and sulfate ions. Residues of DNS3H are labeled in black. The PcrA helicase structure in complex with ADPNP (residues in pink) is overlaid (Velankar et al., 1999). The Manganese molecule coordinated by residues Glu-85 and Asp-284 of DNS3H (motif II) is represented as a grey sphere.

Both the YFV helicase reported by Wu et al., and the DNS3H helicase structure reported here have trapped an enzyme product complex. Domain closure has been suggested to be a general feature of helicase upon NTP binding (Kim et al., 1998; Bernstein et al., 2003; Velankar et al., 1999), however, as mentioned for the YFV helicase, the P-loop adopts the same conformation in the nucleotide complex and in the free enzyme (Wu et al., 2005). A comparison of the six molecules in the three crystals reveals a correlation between the stability of the P-loop and the binding of ligand. In essence, the presence of the Mn$^{2+}$ and sulfate ions stabilizes the conformation of the P-loop, as shown by a lower average temperature factor. Conversely, in the absence of the Mn$^{2+}$ and sulfate ions, the glycine-rich P-loop is more flexible (Fig. 3. 9). The r.m.s deviations among P-loops within different crystals is 0.9 Å, with a maximum deviation of 2.4 Å. The flexibility of the P-loop is reminiscent of the structural changes observed in other helicases upon NTP binding (Bernstein et al., 2003).
Fig. 3. 9 Comparison of P-loops of six molecules in three crystals.

Comparison of distribution of the temperature factors along the α-carbon trace of P-loops of from three crystals. Amino acids of P-loop (Gly-196 to Lys-201) are labelled. The color coding scheme is defined in the bar. The average B factors of different P-loops are shown in top left corner.

As suggested previously (Kim et al., 1998; Story et al., 1992), a number of residues surrounding the NTP binding pocket are likely to function as sensors, by coupling nucleoside 5’ triphosphate binding and hydrolysis with nucleic acid recognition and duplex unwinding, through concerted conformational changes. In this respect, both His-287 (motif II) and Gln-456 (motif VI) might play important roles. In our structure, the imidazole side chain of His-287 forms a hydrogen bond with the carbonyl oxygen of Glu-412 (motif V). It is also at right distances to form polar contacts with the side chains of Glu-285 (motif II), Thr-317 (motif III) and...
Gln-456. In the JEV and HCV NS3 helicase, substitution of this Histidine residue by an Alanine (yielding altered motifs II with sequences DEAA and DECA respectively) dramatically reduces the helicase activity, while most of the NTPase activity is retained (Utama et al., 2000; Heilek et al., 1997). Mutagenesis studies performed on Gln-254 (motif III in PcrA) (Dillingham et al., 1999) which is spatially equivalent to Gln-456 of DNS3H (Fig. 3.8) have demonstrated a correlation between the nature of the charge of this residue and the coupling between ATPases and helicase activities and have led these authors to propose that this residue might act as a “γ-phosphate sensor” (Dillingham et al., 1999; Story et al., 1992). We also note in the same segment of the polypeptide chain (motif VI) the presence of two buried Arg residues: Arg-457 and Arg-458 which point away from the NTP binding site. The strictly conserved residue Arg-457 makes a hydrogen bond with the carbonyl oxygen of residue 427 from domain II. This interaction might play an important role by transmitting conformational changes between the NTP binding site and the long β4A’ β4B’ hairpin, which abuts onto domain III. (Structure-based mutants of DNS3H are described in Sampath et al., 2006.)

3.5.4 Nucleic acid binding sites

The structure of DNS3H reveals a tunnel at its center surrounded by residues emanating from the three domains. This tunnel is lined with a number of basic residues and is wide enough to accommodate a single-stranded nucleic acid substrate of about six nucleotides, but not a duplex (Fig. 3.10). DNS3H contains an unusually high proportion of charged residues and the distribution is
asymmetric (Fig. 3. 10). The face lined by the tunnel bears an excess of positively charged residues, with several basic patches able to accommodate nucleic acid duplexes, while the other face is more negatively charged. This suggests that electrostatic repulsion might play an important role, possibly by propelling the protein along the polymeric substrate and by preventing the reannealing of unwound nucleic acid strands.

Fig. 3. 10 Electrostatic representation of DNS3H.

Surface representation of the DNS3H protein in the same orientation as in Fig. 3. 2 A, illustrating the presence of a tunnel across this face of the protein and the presence of several basic patches able to accommodate nucleic acid substrates (see text). The right panel is a view with the molecule rotated by 180° around a vertical axis. An excess of negative charges is visible on this face of DNS3H. Positive electrostatic potentials are colored blue and negative in red. Putative interaction between DNS3H and a single stranded nucleic acid substrate is shown. The model was obtained by superposition with the HCV helicase in complex with a dU₈ oligonucleotide (PDB code: 1A1V) (Kim et al., 1998) see text. The likely
polarity of the nucleic acid substrate (yellow sticks) is indicated in Fig. 3.11. The right panel is a view with the molecule rotated by 180° around a vertical axis, an excess of negative charges is visible on this face of DNS3H.

In our structure, a sulfate ion, from the crystallization buffer, is located inside the tunnel, next to N-terminal end of helices $\alpha_{1'}$ and $\alpha_{7''}$. It is hydrogen bonded to the $O_\gamma$ of Thr-408 (motif V) and the main chain amide group of Arg-387 (Fig. 3.12). We generated a model for single-stranded nucleic acid binding by superimposing DNS3H to the hepatitis C virus NS3 helicase bound to a deoxyuridylate octamer (dU8) oligonucleotide (Kim et al., 1998) (Fig. 3.11). In this model, the dU8 lies in a groove between the three domains and makes interactions with residues from motifs Ia, IV, and V. This superposition reveals that the distance between the bound sulfate molecule and the putative backbone phosphate of dU5 is only 1.6Å. Interestingly, a similar interaction is establishd by the phosphate group of dU5 which is hydrogen bonded to Arg-393 amide and Thr-411 $O_\gamma$, which are equivalent to residues Arg-387 and Thr-408 in DNS3H, respectively (Fig. 3.12). This suggests that the bound sulfate ion in our structure occupies the position of a phosphate group of a nucleic acid substrate.

In our model, other residues interacting with the phosphodiester backbone would include Arg-225 (motif Ia), Thr-224 (motif Ia), Thr-264, Lys-366 (motif IV), Arg-538 and Arg-599 from domain III. Interestingly, this model places two basic residues of motif VI Arg-457 and Arg-458, which are extremely conserved across the Flaviviridae, more than 10 Å away from the nearest phosphate of the single-stranded nucleic acid ligand, suggesting that these basic residues are not
directly involved in nucleic acid binding.

Fig. 3. 11 Model of DNS3H-dU8 complex.

Putative interactions between DNS3H and a single stranded nucleic acid substrate. The model was obtained by superposition with the HCV helicase in complex with a dU_8 oligonucleotide (yellow sticks) (PDB code: 1AI1V, Kim et al., 1998), see text. The polarity of the nucleic acid substrate is indicated. The right panel is a view with the molecule rotated by 90° around a vertical axis, emphasizing the flatness of the molecule and the central location of the tunnel.

Fig. 3. 12 Comparison of nucleic acid binding sites.
Left panel: Fo-Fc omit map at 3.5 $\sigma$ of a second sulfate ion bound inside the nucleic acid binding tunnel of DNS3H. Hydrogen bonds between sulfate ion and Arg-387 and Thr-408 are shown as yellow dashes. Right panel: Hydrogen bond interaction between phosphate dU5 and Arg-393 and Thr-411 in HCV NS3 helicase and dU8 complex (1A1V) (Kim et al., 1998).

Comparison between the DNS3H and YFV NS3 helicase structures reveals a significant conformational change in motif V and neighboring residues (Fig.3. 5, Fig.3. 13 and Fig.3. 14). In DNS3H, Arg-457 makes a salt bridge, the “RED bridge”, with the side chain Asp-424 and Glu-412 (Fig.3. 13). Other contacts include a salt bridge between Arg-387, Asp-409 and Asp-290 from domain I, and a salt bridge between Lys-388 and Glu-609 from domain III. These two later interdomain interactions bring three domains together to form a “closed” conformation (Fig.3. 5) which in turn favor the binding of a sulfate ion. In contrast, in the YFV helicase structure, Glu-417 (Glu-412 in DNS3H) is no longer “bridged” to Arg-461 (Arg-457 in DNS3H), instead, it makes a hydrogen bond to the amide group of Arg-392 (Arg-387 in DNS3H). This results in a surface charge switch inside the nucleic acid binding site (Fig.3. 14). In DNS3H, the sulfate ion sits comfortably inside a positively charged pocket which is formed by Arg-387 and Thr-308, however, in YFV helicase, the corresponding sulfate binding pocket is highly negative charged (Fig.3. 14 right panel).
Fig. 3.13 Schematic view of local rearrangements in motif V and neighboring residues in Den and YFV helicases.

Left panel: DNS3H residues (green sticks) with the bound sulfate ion. Right panel: YFV helicase residues (yellow sticks) (Wu et al., 2005). Salt bridge and hydrogen bond interactions are shown as green or yellow dash and distance are labeled. Two panels are shown in the same orientation.

Consequently, the negative charged surface no longer favors the binding of sulfate ion and the bound sulfate will be dislodged. The concomitant effect is a local rearrangement of three domains (open conformation). Arg-392 (Arg-387 in DNS3H) no longer makes a salt bridge with Asp-414 (Asp-409 in DNS3H) and Asp-295 (Asp-290 in DNS3H). Lys-393 (Lys-388 in DNS3H) is 16Å away from Glu-614 (Glu-609 in DNS3H). The local rearrangement brings two Arginines, Arg-230 (motif Ia) and Arg-392 (Arg-225 and Arg-387 in DNS3H), close to form a
positive charged surface which could favor the binding of phosphodiester backbone (Fig. 3.14 right panel). This suggests that electrostatic potential might play an important role in the translocation process. Interestingly, the significant change of conformation in motif V observed in domain II is consistent with the intrinsic conformational flexibility of this loop region in molecular dynamic studies of HCV NS3 helicase (Liu et al., 2001).

**Fig. 3.14 Surface representation comparison of the nucleic acid binding site.**

Surface representation comparison of the nucleic acid binding site of domain I and II between DNS3H and YFV helicase. Positive electrostatic potentials are colored blue and negative are colored red. The bound sulfate ion is represented as orange stick. Residues contributing to the charge difference are shown in stick and labeled. Black arrows indicate the different charged patches. Two panels are in the same orientation. (Domain III has been removed from this view for clarity).
3.6 Helicase mechanism

In general, NA unwinding is viewed as a combination of unidirectional translocation and strand separation processes. Both of these processes are fueled by the energy released from NTP hydrolysis. Although the detailed mechanisms of translocation and strand separation are not known, the unwinding process can be described as:

\[
E \rightarrow E + NTP \rightarrow E + NDP + Pi \rightarrow E + NDP \rightarrow E
\]

The helicase bound near the junction of single-stranded and duplex parts of the NA substrate cycles through different NTP ligation states (Levin and Patel, 2003). NTP binding, hydrolysis and product release act as a switch that induces conformational changes on the helicase NA binding site, forcing it to change its affinity for the NA or to perform a power stroke. These conformational changes drive the unidirectional translocation and unwinding in a stepwise fashion. Several mechanisms of NTPase-coupled translocation and unwinding have been proposed for different helicases (Kim et al., 1998; Korolev et al., 1997; Velankar et al., 1999; Cho et al., 1998; Singleton et al., 2000; Li et al., 2003; Gai et al., 2004). Some of the differences in the proposed mechanisms reflect the diverse biochemical properties of different helicases, such as oligomeric state, the interactions with ssNA versus duplex NA, and the effect of NTP ligation state on the NA binding properties. Considering that helicases show a significant degree of structural homology, it would be desirable to arrive at a general model that, with minor variations, can explain the properties of all helicases. It should be adaptable to different classes of helicases, explain translocation directionality, and helicase function (Levin and Patel, 2003).
Macroscopically observed enzymatic reactions are composed of a large number of microscopic stochastic events. Biological systems have evolved to utilize this randomness to achieve a non-random directed result. It is likely, that helicases as well as other molecular motors are no exception. It has been shown that Brownian motion can be integrated into a model of a molecular motor leading to a simpler, more flexible, and more robust design (Astumian, 2000, 2001; Huxley, 1998; Oster, 2002).

*Fig.3. 15* A and B, show, respectively, the mechanics and energetics of the Brownian motor mechanism. The binding of NS3h to ssDNA is a high energy event. Some special properties of ssNA-helicase interface can utilize this energy for unidirectional translocation by creating a local binding free energy gradient along the ssNA length. This is represented by the sawtooth binding energy profile in *Fig.3. 15* B (*solid line*). In the absence of ATP, helicase is trapped on the NA in its lowest energy state, unable to translocate (*Fig.3. 15* A and B, positions 3 and 5). The binding of ATP switches helicase to a different conformation with a weaker affinity for NA (*Fig.3. 15* A and B, positions 2 and 4). In this weak binding state, the binding free energy is constant along the NA length (*Fig.3. 15* B, *dotted line*), which allows the helicase to slide along the length of the NA in either direction because of Brownian motion (*Fig.3. 15* A and B, position 4). In the weakly bound state, there is a possibility that helicase completely dissociates from the NA and that it is a nonproductive event. The random movement of the helicase along the length of the nucleic acid lasts only a short time because of rapid ATP hydrolysis. After ATP is hydrolyzed and products are released, the helicase rebinds NA tightly. Had the Brownian movement led to a back-walking of the helicase, then upon
rebinding to the nucleic acid the helicase will end up in the same starting position, without net movement (Fig. 3.15 B, position 3). The helicase that diffuses in the forward direction (Fig. 3.15 B, position 5) will end up one step forward from its original position upon tight binding to the nucleic acid. Thus repeated binding and release of the helicase, catalyzed by the NTPase cycles results in a net unidirectional translocation of the helicase along the length of the NA (Levin and Patel, 2003).

![Diagram of Brownian motor mechanism of unidirectional translocation.](image)

The mechanism is illustrated in terms of its possible mechanics (A) and energetics (B). The helicase in A is represented by the shaded area. The helicase-NA binding free energy is plotted along the length of the nucleic acid in B. (See text)
The BM model for monomeric helicases predicts limited processivity of translocation and unwinding *in vitro* because the helicase holds on to the ssNA only via one binding site, and therefore it is more likely to dissociate during the sliding phase. Experimental observations support this prediction (Ali and Lohman, 1997; Dillingham *et al.*, 2000). A dimeric or hexameric helicase with two or six ssNA binding sites can implement a BM mechanism with a greater efficiency. This is especially true if the helicase subunits coordinate their ATPase cycles, and take turn binding and hydrolyzing ATP. The idea of BM or molecular ratchet model is well established in the field of ion pumps and filament motors, while only a few researchers consider a BM mechanism for helicases (Doering *et al.*, 1995). Therefore the BM translocation model is consistent with most known facts about helicases. The BM model is preferred as a simpler, flexible, and a more evolvable model (Levin and Patel, 2003).

### 3.7 Helicase mechanism of DNS3H

Differences between the Dengue virus and YFV NS3 helicase structures, especially the charge switch at the surface of the ss nucleic acid binding site, allow us to make the following hypothesis. The two different structures might represent two crucial states in the Brownian motor model, the tightly bound state and weakly bound state. In essence, in the absence of ATP, DNS3H adopts a closed conformation (*Fig.3. 5, red*), by burying Glu-412, which makes a salt bridge with Arg-457, inside domain II, and anchoring the three domains together, mainly through contacts between R-387 and D-290 (domain I), K-388 and E609 (domain
III), Lys-366 and Ser-602 (domain III). This arrangement leaves the positive surface exposed which in turn favors the binding of phosphodiester backbone of nucleic acid and therefore the tightly bound state. Upon ATP binding, the helicase adopts an open conformation, which is represented by the YFV structure (Fig. 3.5, blue), by breaking the contacts described above, the protruding side chain of Glu-412 is able to dislodge the phosphodiester backbone of nucleic acid and therefore induce the weakly bound state. In this transient state, DNS3H slides randomly and bidirectionally along the single stranded nucleic acid. Interaction between the complementary strand and helicase exerts extra constraints to the movement of helicase, which in turn favors the slippage along a 3’ to 5’ direction. ATP hydrolysis makes a power stroke and helicase rebinds tightly to nucleic acid. This results in a net translocation along the 3’ to 5’ direction. Interestingly, our model puts the duplex in contact with several basic residues protruding from domain II: Arg-432, Lys-366, Arg-376 and Lys-377 (helix α1’), Lys-381, Lys-396 and Arg-398 (helix α2’); and Lys-418 and pointing towards the N-terminal domain, where the NS3 protease domain could provide additional interactions.

One implication of our model is that a 3’ single-stranded tail of a minimum of about 8 to 10 nucleotides would be required for initiation of the unwinding reaction. This requirement is in agreement with our preliminary observation of the absence of unwinding activity of DNS3H on a blunt-ended duplex DNA substrate (not shown). The origin of the 3’ to 5’ directionality (as opposed to 5’ to 3’), however, is more difficult to interpret. Detailed mechanistic studies as well as experimental structures of complexes with nucleic acid substrates are required to resolve these issues.
3.8 The full length NS3 protein

Flaviviruses full length NS3 have been shown to exhibit greater helicase activity as compared to the helicase domain alone (Yon et al., 2005). The dispensability of an active catalytic triad for stimulatory effect suggests that the overall folding of the N-terminal protease domain contributes to this enhancement (Yon et al., 2005). The crystal structure of an engineered HCV protein containing the complete NS3 and the protease activation fragment from the non-structural protein 4A (NS4A) reveals that helicase and protease domains form separate domains connected by a single β-strand, with the helicase domain exposed to the solvent (PDB: 1CU1; Yao et al., 1999). The protease domain packs behind the helicase domain and accommodates its cofactor NS4A, which extends from the C-terminal end of the helicase domain into its catalytic site. Even though other domain organizations cannot be ruled out, the packing scheme employed by HCV NS3 might be generally applicable to flaviviruses NS3. A model of dengue full length NS3 was generated based on superposition of structures of dengue NS3 helicase and protease (PDB: 2FOM; Erbel et al., 2006) domains to HCV NS3 (PDB: 1CU1) helicase and protease domains, respectively. The resultant model places the C-terminal end of protease domain 59 Å away of the N-terminal end of helicase domain (Fig. 3.16). However, in light of the flexibility of the N-terminal residues of DNS3H helicase, which adopts an extended conformation and are involved in crystal contacts with neighboring molecule, a considerable local rearrangement is possible which would bring the protease domain and helicase domain together. Nevertheless, the ensuing model does not provide significant information concerning the stimulatory effect on helicase activity. Further structure studies of flaviviruses full length NS3 and in complex with nucleic acids are thus desirable.
Fig. 3. 16 Model of the dengue full length NS3 protein.

The model is generated based on superpositions of structures of Dengue NS3 helicase and protease (PDB: 2FOM; Erbel et al., 2006) domains to 1CU1 (Yao et al., 1999) helicase and protease domains, respectively. Dengue NS3 helicase domain starts from amino acid 171 to 618, with three extra amino acids (Ala-Met-Ala) introduced by NcoI at the N-terminal. Dengue NS3 protease domain starts from 18 to 167 (Erbel et al., 2006). The resulting distance between the N-terminal end of helicase domain and the C-terminal of protease domain is about 59 Å. Dengue NS3 helicase, protease and NS2B cofactor are colored in blue, purple and yellow, respectively. The catalytic triad (His-51, Asp-75, Ser-135) is depicted as cyan sticks.
3.8 The flavivirus replication complex

The flavivirus NS3 helicase is a member of a large macromolecular complex of the viral proteins, consisting of viral protease cofactor NS2B, methyltransferase and RNA-dependent-RNA-polymerase NS5, and perhaps other viral nonstructural proteins and host cell proteins. It is responsible for the replication of the genome, processing of polyprotein and initiation of genome packaging. The highly conserved 3’ untranslated region of the viral genome is also involved in replication complex formation (Chen et al., 1997). The precise nature of the flavivirus replication complex remains elusive but NS3 and NS5 are thought to have crucial contributions.

Interactions between NS3 and NS5 were detected by immunoprecipitation both in Dengue and Japanese encephalitis virus (Chen et al., 1997; Kapoor et al., 1995). The physical contact region was well characterized and mapped to the C-terminal 300 residues of NS3, corresponding to domain II and III of NS3 helicase (Johansson et al., 2001), and to 40 residues in the nuclear localization signal (NLS) region of NS5 (Johansson et al., 2001), which belongs to the NS5 RdRp domain (Yap et al., to be published). NS3 competes with importin-β for the same NLS binding site of NS5 to regulate NS5 trafficking, remaining in cytoplasm or entering into nucleus (Johansson et al., 2001). Furthermore, NS5 stimulates NS3 NTPase and RTPase activities (Yon et al., 2005).

Interestingly, the interaction between HCV NS3 and NS5 is of different nature. HCV RNA-dependent-RNA-polymerase does not have a nuclear localization signal site, and is not found inside the nuclei. NS5 interacts with NS3 through the
N-terminal protease domain but not with the C-terminal helicase domain (Ishido et al., 1998). Structural data of NS3 and NS5 complex are required to elucidate the details of their interaction.

3.9 Conclusion and perspectives

In conclusion, we reported the three-dimensional structure of the Dengue virus NTPase/helicase catalytic domain. The structure reveals a tunnel at one face that is large enough to accommodate single-stranded RNA. The bound manganese and sulfate ions inside NTPase binding pocket reveals residues involved in the divalent metal-dependent NTPase catalytic mechanism. Structural difference between yellow fever virus NS3 helicase (Wu et al., 2005) and that of dengue, especially the rigid body movement in domain 3 and the conformational change in motif V, provide evidence for the “Brownian motor” mechanism, in which the unidirectional translocation is fueled by single-stranded DNA binding while ATP binding allows for a brief period of random movement that prepares the helicase for the next cycle. The possible NA binding model proposed reveals potential residues implicated in NA-protein interaction, providing an experimental basis to carry out site-directed mutagenesis to probe the function of charged residues for NA binding and duplex unwinding (Sampath et al., 2006).

Some mechanistic properties of dengue NS3 helicase remain unclear. Detailed mechanistic studies, such as single molecule experiment, as well as structures of complexes with nucleic acid substrates are required to unravel these issues. Dengue NS3 helicase domain alone only displays basal NTPase and helicase
activities; optimum activities requires the presence of the N-terminal protease domain. Furthermore, NS5 has been shown to interact with NS3 (Chen et al., 1997; Kapoor et al., 1995) and stimulate the NS3 NTPase and RTPase activities in an equimolar manner (Yon et al., 2005). Structural data of the full length NS3 as well as the NS3 and NS5 complex are required to elucidate the details of their interaction. The present studies should enhance our understanding of the properties of dengue NS3 protein and contribute to the development of specific antiviral compounds.
REFERENCES


References

731-737.


---

**References**


References


References

dengue virus type 2 using vaccinia virus. Dimerisation of the NS1 glycoprotein. 


References


DEAH box splicing factor Prp22 are important for RNA unwinding, and helicase-defective Prp22 mutants are suppressed by Prp8. *J Biol Chem.* **279**, 8617-8626.


References


Wu, S.J., Grouard-Vogel, G., Sun, W., Mascola, J.R., Brachtel, E., Putvatana, R.,


Young, P.R., Hilditch, P.A., Bletchly, C., and Halloran, W. (2000). An antigen capture enzyme-linked immunosorbent assay reveals high levels of the dengue


APPENDIX

MEDIUM AND SOLUTION

1) Buffer A
20 mM Na$_3$PO$_4$, pH 7.4, 0.5 M NaCl, 40 mM imidazole

2) Buffer B
20 mM Na$_3$PO$_4$, pH 7.4, 0.5 M NaCl, 40 mM imidazole

3) Buffer C
20 mM Tris-HCl, pH 8.0

4) Buffer D
20 mM Tris-HCl, pH 8.0, 1M NaCl

5) Buffer E
20 mM Tris-HCl, pH 8.0, 50 mM NaCl, 2 mM CaCl$_2$

6) Buffer F
20 mM Na$_3$PO$_4$, pH 7.4, 0.5 M NaCl

7) Buffer G
20 mM Tris-HCl, pH 7.4, 0.5 M NaCl, 1 mM EDTA, 1 mM DTT

8) Buffer H
10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 1 mM DTT

9) M9 minimum medium
To 750 ml sterile deionized H$_2$O, add 6 g Na$_2$HPO$_4$; 3 g KH$_2$PO$_4$; 1 g NH$_4$Cl; and 0.5 g NaCl. Autoclave.