Functional and structural studies of the flavivirus RNA-dependent RNA polymerase

Yap, Thai Leong

2008

Yap, T. L. (2008). Functional and structural studies of the flavivirus RNA-dependent RNA polymerase. Doctoral thesis, Nanyang Technological University, Singapore.

https://hdl.handle.net/10356/18856

https://doi.org/10.32657/10356/18856

FUNCTIONAL AND STRUCTURAL STUDIES OF THE FLAVIVIRUS RNA-DEPENDENT RNA POLYMERASE

YAP THAI LEONG

School of Biological Sciences

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

2008

ACKNOWLEDGEMENT

I want to thank everyone who has contributed to this thesis, and all those that have made these years at Lescar Laboratory at Nanyang Technological University (NTU) and at the Dengue Unit of Novartis Institute for Tropical Diseases (NITD) so enjoyable.

I am especially grateful to:

Dr. **Julien Lescar** for accepting me as a PhD student. Deeply appreciated for his continuous support, guidance and giving me freedom to do things during all these years.

Dr. **Subhash Vasudevan** for being my co-supervisor and giving me opportunity to do part of my project in your laboratory, as well as sponsored me all the useful training courses. The financial support from NITD has been greatly appreciated. Your encouragement, enthusiasm, knowledge and patience led me to make progress. Dr. **Alex Matter** and Dr. **Thomas Keller**, for their support and giving me opportunity to work on NS5 project with the drug discovery team at NITD during these years of fruitful collaboration.

Dr. **Chen Yen Liang** for supporting me with your knowledge about the flavivirus NS5 enzyme during my stay at NITD. Also for his kindness and helpfulness when written my manuscripts and thesis. Dr. **Xu Ting** for your knowledge and guidance in protein crystallography. You always have been willing to share this with me and Dahai. Dr. **Celine Nkenfou** for all your wonderful sharing and we have worked together for the Fab antibody project. Without you, the antibody project would have been impossible. Dr. **Lim Siew Pheng**, for your patience and guidance in the MTase project. **Samantha** and **Joanne** for the patience

and effort to work together to explore the NS5 project. Former colleague **Wan Yen** for her helpfulness and all things that you have taught me.

My close collaborators at the Dengue Unit. **Dahai**, **Chrisophe Bodenreider**, **Nood**, **Wai Yee**, **Daying**, **Liu Wei** and **Indira** for all the support and scientific discussion we have had during the years. I can't imagine what my time at the institute would have been without you. **Hao Ying**, **Katja Fink**, **Swee Hoe**, **Andy**, **Alex**, Dr. **Gu Feng**, Dr. **Wong Qing Yin**, Bo ping, **Cedric** and Dr. **Wouter** for their support and the discussions I have had with you. Without you, the laboratory would not have been so enjoyable.

My colleagues at the Lescar laboratory. Dr **Masayo Kotaka**, for all discussions that we had and worked together at synchrotrons. **Dahai** for keeping the X-ray machine running. **Ravi**, **Hooi Chen**, **Yee Hwa**, **Mitchell** and **Dahai** for all the support I have got from you or whenever I have had problem that I needed advice or help. **May Chong** from NTU, **Azni**, **Stephanie**, **Cindy** and **Eric** from NITD for your support with all administrate details.

Dr **Clemens Schefler**, Dr. **Sandra Jacob** and Dr **Hans Widmer** for help with data collection for this work that was performed at Beamline X10SA (PXII) of the Swiss Light Source, Paul Scherrer Institute, Villigen, Switzerland. Dr. **Bruno Canard** and his team at Marseille for the past years collaboration and communication.

Finally, I want to give thousands of roses to my family for all the understanding and support that I have received from you during these years.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PUBLICATIONS

Publication related to this thesis:

Yap,T.L[#]., Xu,T[#]., Chen,Y.L., Malet,H., Egloff,M.P., Canard,B., Vasudevan,S.G., and Lescar,J. (2007). **Crystal Structure of the Dengue Virus RNA-Dependent RNA Polymerase Catalytic Domain at 1.85-Angstrom Resolution**. J. Virol. *81*, 4753-4765.

Yap,T.L., Chen,Y.L., Xu,T., Wen,D., Vasudevan,S.G., and Lescar,J. (2007). **A multi-step strategy to obtain crystals of the dengue virus RNA-dependent RNA polymerase that diffract to high resolution**. Acta Crystallograph. Sect. F. Struct. Biol. Cryst. Commun. *63*, 78-83.

Other publications:

Kroschewski,H., Lim,S.P., Butcher,R.E., **Yap,T.L.**, Wright,P.J., Lescar,J., Vasudevan,S., and Davidson,A.D. (2008). **Mutagenesis of the Dengue virus type 2 NS5 methyltransferase domain**. J Bio Chem. *28*, 19410-19421.

Malet,H., Masse,N., Selisko,B., Romette,JL., Alvarez, K., Guillemot,JC., Tolou,H., **Yap,TL.**, Vasudevan,S., Lescar,J., Canard,B. (2008). **The Flavivirus polymerase as a target for drug discovery**. Antiviral Res. *80 (1)*, 23-35.

Lim, S.P., Wen, D., **Yap, T.L.**, Chung, K.Y., Lescar, J., Vasudevan, S. (2008). **A scintillation proximity assay for dengue virus NS5 2'-O methyltransferase- kinetic and inihibition analyses**. Accepted.

Niyomrattanakit,P., Chen,Y.L., **Yap,T.L.**, Joanne,L.Y.H., Lee,WY., Brooks,A., Lescar,J., and Vasudevan,S. (2008). **Scintillation proximity assay and biochemical characterization of a recombinant dengue virus NS5 RNA-dependent RNA polymerase expressed from insect**. Manuscript in preparation.

CONFERENCE AND TRAINING COURSE

Conferences/ Symposium

1. Symposium on Dengue Fever and Tuberculosis. (Jul 2004). *Singapore.*
Host by Novartis Institute for Tropical Diseases.

2. 2nd Asian Regional Dengue Research Network Meeting. Strategies for Vaccines and Therapeutics for Dengue and other Flaviviral Diseases. (Sep 28-30, 2005). *Singapore.*
***Poster presentation*:** Expression, Purification, and Biochemical Characterization of NS5 Dengue Virus Type-2 from Baculovirus system.

3. Joint 3rd AOHUPO and Fourth Structural Biology and Functional Genomics Conference. (Dec 4 -7, 2006). *Singapore.*

4. 3rd Asian Regional Dengue Research Network Meeting. (Aug 22-24, 2007). Taipei, *Taiwan*. Scholarship award.
Colloquium presentation: Crystal structure of the Dengue Virus RNA-dependent-RNA polymerase Catalytic Domain.

Training courses

1. Training Workshop on X-ray Crystallography, in conjunction 3rd international conference on structural and functional genomics. (Nov 17-21, 2004). *Singapore.*

2. 2nd Experimental Course in Antibody Phage Technology. (Feb 13 -17, 2006). ETH Zurich, *Switzerland.*

3. Hercules (Higher European Research Course for Users of Large Experimental Systems) Specialized Course 6. Advances and New Applications of Synchrotron Radiation for Structural Biology. (7-13th Oct 2007). Grenoble, *France.*

ABBREVIATIONS

3'dGTP: 3'deoxyguanosine triphosphate

$3D_{pol}$: 3D poliovirus

α: alpha

β: beta

γ: gamma

aa: amino acid

β-ME: β-mercaptoethanol

BVDV: bovine viral diarrhea virus

CS: complementary sequence

DdDp: DNA-dependent DNA polymerase

DdRp: DNA dependent RNA polymerase

DENV: dengue virus

DF: dengue fever

DHF: dengue hemorrhagic fever

DMSO:dimethyl sulfoxide

dsRNA: double strand RNA

DSS: dengue shock syndrome

DTT: dithiothreitol

E: enzyme

EDTA: ethylenediaminetetraacetic acid

ER: endoplasmic reticulum

FL: full length

FMDV: Foot Mouth Disease Virus

h: hour

HBV: hepatitis B virus

HCV: hepatitis C virus

HIV-RT: human immunodeficiency reverse transcriptase

IRES: internal ribosomal entry site

JEV: Japanese encephalitis virus

K: kelvin

K_d: binding affinity

KF: klenow Fragment

KUNV: Kunjin virus

MTase: methyl transferase

Min: minute

Mg: magnesium

Mn: manganese

MW: molecular weight

NC: nucleocapsid core

NGC: New Guinea C

NI: nucleoside inhibitors

NLS: nuclear localization signal

NNI: non-nucleoside inhibitors

NS: non structural

PEG: polyethylene glycol

PPi: pyrophosphate

p/t: Primer/template

RdDp: RNA dependent DNA polymerase

RdRp: RNA dependent RNA polymerase

RF: replication form

RI: replicative intermediate

r.m.s. deviation: root mean square deviation

RT: room temperature

RTPase: RNA triphosphatase

S: substrate

SAM: S-adenosyl-methionine methyltransferase

SDS-PAGE: Sodium dodecyl sulphate-polyacrylamide gel electrophoresis

SL: stem loop

SLA: stem loop A

TBEV: tick-borne encephalitis virus

tp: triphosphate

TGN: Tran Golgi network

TSVO1: Townville O1

UTR: untranslated regions

VP: vesicle packets

WNV: west nile virus

YFV: yellow fever virus

Zn: zinc

ABSTRACT

Dengue fever, a neglected emerging disease for which no vaccine or antiviral agents exist at present, is caused by dengue virus, a member of the Flavivirus genus. The NS5 protein from dengue virus (DENV) is bifunctional and contains 900 amino acids. The S-adenosyl methionine transferase activity resides within its N-terminal domain, and residues 270 to 900 form the RNA-dependent RNA polymerase (RdRp) catalytic domain. Viral replication begins with the synthesis of minus-strand RNA from the dengue virus positive-strand RNA genome, which is subsequently used as a template for synthesizing additional plus-strand RNA genomes. This essential function is catalyzed by the NS5 RdRp and thus the protein represents an interesting target for the development of specific antiviral compounds. Therefore, this work set out to investigate the structural basis of DENV RdRp in order to provide clues to understand the molecular details of its role and function in viral genome replication. This thesis work presents three aspects of study that consists: (1) the protein expression strategy and characterization of NS5 polymerase, (2) a multi-step strategy to obtain crystal of DENV RdRp for high resolution diffraction and soaking studies and (3) structural insight of the DENV RdRp. To obtain soluble proteins for structural and functional studies, an extensive array of approaches has been explored, which comprises (1) expression hosts and conditions, (2) refolding, (3) limited proteolysis and (4) homologous protein screening. The enzymatic activity shows that the RdRp domain is less active than the full length (FL) NS5, which is also supported insights from the limited proteolysis experiments and Fab epitope mapping. The limited proteolysis results indicate that the catalytic domain RdRp protein adopts a more open and flexible conformation than the FLNS5. Antibody Fab fragments that were obtained in this study and found to recognize conformational epitope interacted more weakly with the RdRp domain compared with the FLNS5. Together these

two approaches imply conformational difference between the RdRp and FLNS5 that possibly account for the difference in activity. For structural studies, a very intensive campaign to crystallize the FL protein has been attempted in this work using proteins expressed from both insect baculovirus expression system and *E.coli* expression. NS5 from two different strains belonging to serotype 2 and WNV were investigated. Complex of NS5 with Fab fragment and also NS3 were investigated for crystal production. In parallel, truncated constructs expressing various parts of NS5 were also exhaustively studied with the information gleaned from limited proteolysis. These studies led to the first successful crystallization of a DENV RdRp from serotype 3. Initially the crystals diffracted poorly and were markedly improved by the addition of divalent metal ions and air dehydration, which led to crystals with high resolution diffraction (1.85 Å) thus making it possible to use molecular replacement to solve the structure.

The structural basis of DENV 3 RdRp shed light on the role and function of the RdRp. The architecture of the protein assumes the canonical right-hand conformation consisting of fingers, palm and thumb, which is characteristic of known polymerase structures. The NS5 nuclear localization sequences (NLS), previously thought to fold into a separate domain, form an integral part of the RdRp subdomains. The structure also reveals the presence of two zinc ion binding motifs in the fingers and thumb subdomains respectively. In the absence of a template strand, a chain-terminating nucleoside analogue binds to the priming loop site. Residues at this site that coordinate the tri-phosphate moiety play critical role in de novo initiation. Finally, the roles of metal ion in catalysis and fidelity model are highlighted. These results should inform and accelerate the structure-based design of antiviral compounds against dengue virus.

CHAPTER 1

INTRODUCTION

1.0 Dengue- an Emerging Arboviral Disease

In the recent decades many mosquito-borne viruses have became medically important pathogens that cause diseases in both humans and domestic animals. These diseases have become broadly classified as emerging infectious diseases (Gubler, 2002b; Mackenzie et al., 2004). Dengue is a mosquito-borne arboviral (arthropod borne virus) disease, which causes a wide spectrum of disease symptoms that ranges from an undifferentiated fever to self limiting dengue fever (DF), or progress to the potentially fatal dengue hemorrhagic fever (DHF) and dengue shock syndrome (DSS).

The term of arbovirus is an ecological representation that describes viruses which require blood sucking arthropod for transmission between hosts (Calisher et al., 1981). Arboviruses are diverse and can be classified into 8 viral families and 14 genera. Most arboviruses of public heath importance belong to three families: Flaviviridae, Togaviridae, and Bunyaviradae (Gubler, 2001). Flaviviridae are enveloped viruses with positive-strand RNA genomes that can be grouped into three genera, hepacivirus (e.g, hepatitis C virus, HCV), pestivirus (e.g, bovine viral diarrhea virus, BVDV) and flavivirus. Within the genus, the viruses can be further subdivided into antigenic complexes according to serological criteria, or into clusters, clades and species on the basis of molecular phylogenetics. Flavivirus genus contains over 70 viruses, of these ~ 40 are mosquito-borne, 16 are tick-borne and around 18 are classified in the no known vector group (Fig 1.2). Several members of the Flavivirus genus, e.g., dengue virus (DENV), yellow fever virus (YFV) Japanese encephalitis virus

(JEV), tick-borne encephalitis virus (TBEV) and West Nile virus (WNV) are well known medically important pathogens that afflict humans (Fig 1.2).

The demographic and societal changes such as population growth, unplanned urbanization, increase movement of viruses in infected humans through modern transportation and the lack of mosquito control account for the increase of epidemic activity and colonization of new areas (Mackenzie et al., 2004). In addition, other natural factors such as genetic change in the virus, host-vector relationship, bird migration and climate changes may have contributed to virus spread. It is now estimated that 2.5-3 billion of the world's population are at risk of dengue infection (Fig 1.1). Annually, around 50-100 millions cases of DF occur world-wide and of these about ~500 000 patients develop the more severe disease that lead to hospitalizations and approximately 20 000 deaths, mainly in children (Gubler and Meltzer, 1999; Gubler, 2002a; Guzman and Kouri, 2002). Although much lower in terms of mortality, the economic impact is similar to that of other major infectious diseases such as malaria (Boutayeb, 2006; Clark et al., 2005).

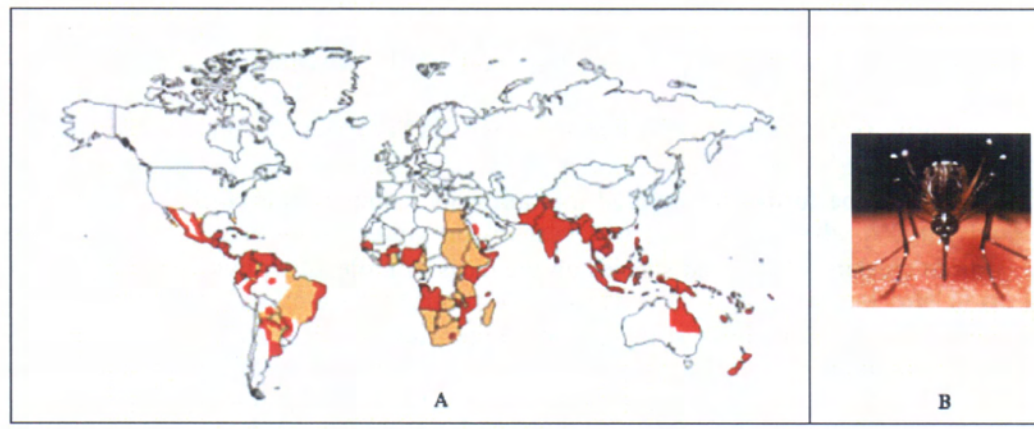


Fig 1.1.Distribution of Dengue and its predominant vector *Aedes aegypti*

(A) World distribution map of dengue and *Aedes aegypti* in 2005. Orange represents areas infested with *A. aegypti* and red represents areas with *A. aegypti* and dengue epidemic activity. (Adapted from CDC). (B) The epidemic vector *Aedes aegypti* showing the characteristic striated legs.

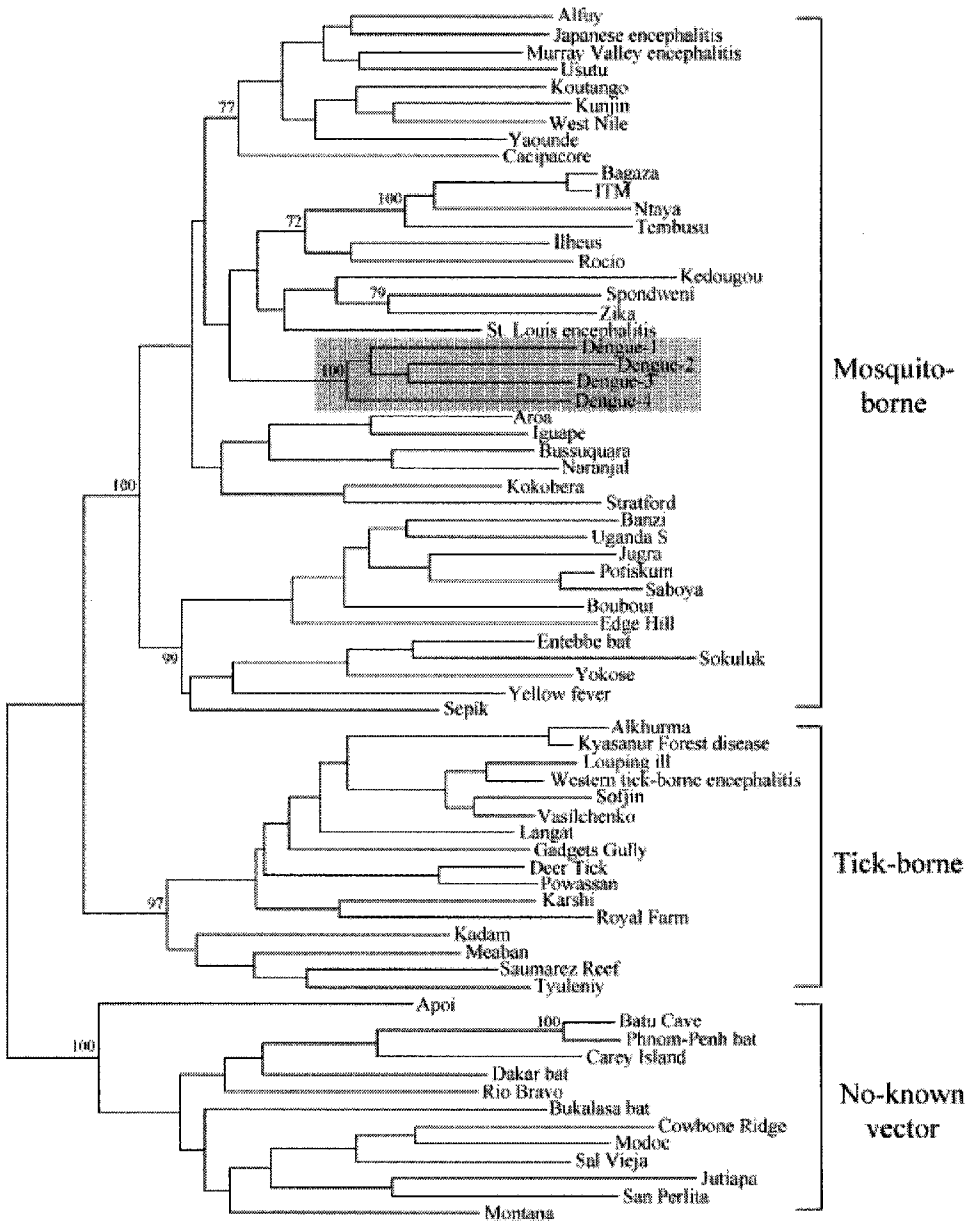


Fig 1.2.Maximum likelihood phylogeny of flavivirus

Maximum likelihood (ML) phylogeny of 70 viruses from flavivirus based on the NS5 gene sequence, classified into tick borne, mosquito borne and no-known vector group. DENV are that classified under Mosquito borne is highlighted. Adapted from Holmes et al., 2003.

1.1 Epidemiology of Dengue virus

As alluded to briefly, transmission of flaviviruses transmission can be sub-grouped into three ecological groups- mosquito borne, tick borne and no-known vector (Fig 1.2). Most flaviviruses are zoonoses that depend on animal hosts for their maintenance in nature. Humans are usually the dead-end or incidental hosts that do not contribute to the transmission cycle. However, DENV have adapted to humans for maintenance in the urban environment (human-mosquito-human transmission cycle). A sylvatic cycle in animal reservoirs involving mosquito-monkey-mosquito transmission cycle are maintained in the jungles of Africa and South East Asia (Gubler, 1996; Gubler, 2001). DENV consists of four antigenically related virus serotypes called DENV-1, DENV-2, DENV-3 and DENV-4 (Zanotto et al., 1996). Although extensive cross reactivity exist among the viruses in serological tests, there is no cross protection in humans. A person living in an endemic area can have as many as four infections, one with each serotype during their life span. The DENV are transmitted among humans by infective female mosquitoes of the genus *Aedes*. The most efficient dengue epidemic vector is *Aedes aegypti*. DENV1 was first isolated by Japanese and American investigators in the Pacific during World War II, respectively (Hotta, 1952; Sabin, 1952). Later, DENV2 was isolated by Sabin. DENV3 and DENV4 were subsequently isolated in an epidemic from Philippines and Thailand respectively (Hammon et al., 1960). Since then, no new serotypes have been documented although the popular press has described serious outbreaks in Indonesia and other parts of the world as being possibly due to new serotypes. Dengue epidemics occurred every 10-30 years until World War II. However, after World

War II, movement of troops and war materials presented ideal conditions for the spread of DENV, mosquito and circulation of multiple DENV serotypes (hyperendemicity) in Asian countries (Henchal and Putnak, 1990; Kautner et al., 1997). The first DHF outbreaks occurred in Philippine in 1950s and dengue epidemic spread throughout Southeast Asia within 20 years (Hammon et al., 1960). The disease remained localized in south-east Asia throughout 1970s and epidemics occured every 3 to 5 years in hyperendemic areas. By mid 1970s, DHF had become a leading cause of hospitalization and death among children in the Southeast Asia. In 1980s and 1990s, epidemic DHF spread west to India, Pakistan, Sri Lanka, the Maldives and east into China from Southeast Asia (Gubler, 1998).

1.2 RNA Virus Replication

The genome of all organisms consists of DNA or RNA and is replicated by enzymes called polymerases. The process of copying a parental RNA to RNA new daughter RNA is mediated by enzymes called RNA dependent RNA polymerases (RdRp). This class of enzyme plays an important role in the evolution and maintenance of a large number of RNA viruses. Viral RNA replication is a highly error prone process, where typically errors in the order of 10^{-4} are observed (Twiddy et al., 2003). The error-prone mechanism is not only crucial for replication and but it also serves as an engines to drive viral evolution and genome variability (Crotty et al., 2001). The RdRp enzymes have crucial functions in genome replication, mRNA synthesis, RNA recombination and gene silencing for copying RNA templates into RNA products. RNA viruses are divided into three major classes which are differentiated by whether the infectious virion particles contain a genome as double strand RNA (dsRNA), positive strand RNA (+) or negative strand RNA (-). The general

characteristic of the replication of these three major classes are discussed below and is summarized diagrammatically in Fig 1.3.

1.2.1 Positive strand RNA viruses

Positive strand RNA viruses form the largest class of RNA viruses. Dengue virus is a positive-strand RNA virus and its genome is directly infectious, unlike the (-) or dsRNA viruses. The (+) RNA viruses replicate in the cytoplasm of infected cells within membrane visicles in a multiprotein complex. Following uncoating, the (+) strand is first translated by host machinery proteins into viral proteins and then it serves as template for new genomes replication via a (-) strand RNA that forms an intermediate dsRNA (Fig 1.3). Viral RdRps act together with other viral and host factors in viral replication. They act as a complex to (1) select RNA template site initiation and RNA synthesis, (2) differentiate RNA genomic replication from mRNA transcription, (3) modify RNA products with 5' caps or 3' polyadenylation and other functions (Lai, 1998). When examined *in vitro*, most RdRps can act alone on RNA templates and they display poor substrate/template specificity, that do not match those observed *in vivo* (Kao et al., 2001). For example, HCV RdRp has been reported to use copy back mechanism in vitro but most likely this protein uses de novo initiation *in vivo* (Kao et al., 2001). Because of the relative small genome size, many (+) RNA viruses use both mature and precursor polypeptides or a single protein plays more than one role for RNA replication (Bedard and Semler, 2004), including the flaviviruses. For example, the prM glycoprotein precursor of dengue could acts as chaperone for the proper folding of E protein to form heterodimers efficiciently during infection (see Table 1.1) (Lorenz et al., 2002). In addition, NS3 protein that contains protease, helicase and NTPase enzymatic activities play important roles in viral polyprotein processing and viral replication.

1.2.2 Negative strand RNA Viruses

The most well studied (-) strand RNA virus replication is documented for influenza virus. Transcription is the first step for (-) strand RNA viruses. The RdRp that is present in the virus particle is transferred together with the genome into the infected cell, in the form of a ribonucleoprotein complex (RNP) (Fig 1.3). Concomitantly, RNA replication proceeds with a (+) strand replication intermediate. The RdRp of (-) strand RNA virus is usually a very large protein (250 kDa).The naked (-) strand RNA virus genome requires a RNP and associates with viral structural polyproteins (NP) to function as a template. The NP proteins could function to melt the secondary structure of the RNA genome so that the bases are exposed to be read out by RdRp without dissociation from the RNP and also to encapsidate the newly synthesized RNA into RNP (Baudin et al., 1994). The RNP is usually in helical conformation due to NP-NP interaction (Ruigrok and Baudin, 1995). For influenza virus, the specific binding of the RdRp to the terminal RNA sequences induces a circular or superhelical conformation. In addition, in (-) strand RNA viruses, the phosphoprotein (P) is pivotal for RNA replication. It interacts with the polymerase and presents in the forms of tetramers to mediate its binding to RNP template (Tarbouriech et al., 2000). Compared to (+) RNA viruses, structural information is limited for (-) strand RNA viruses. A model for polymerase complex of influenza virus has been reconstructed based on data from cryo-electron microscopy (Area et al., 2004).

1.2.3 Double stranded RNA viruses

Rotavirus is one of the most well known dsRNA virus. Generally, the genome of dsRNA viruses can be monopartite or divided into fragments (Fig 1.3). The infectious rotavirus particle is made up of three concentric layers of protein and contains a genome consisting of segmented double stranded RNA. Upon infection, RNA polymerases associated with double-

layered virus particle are activated to forms double layer particles (DLP), which serves as transcriptase, resulting in genome transcription (Patton et al., 2004). The capped mRNAs are released into cytoplasm through the tunnel formed between DLPs, which serves as template for new dsRNA synthesis. The dsRNA synthesis is an event that occurs following the gene-specific packaging of viral mRNA into core-like assembly intermediates. Both the transcription and synthesis of new dsRNA is performed in the confined capsid (see Fig 1.3c) to avoid cellular nucleases degrade RNA genome and prevent antiviral responses in the host cells due to the increased copy numbers of dsRNA, before it could be released into infected host cells (Jayaram et al., 2004). Several RdRp structures of dsRNA viruses have been determined, including that of Φ6 RdRp that has been captured as initiation complex.

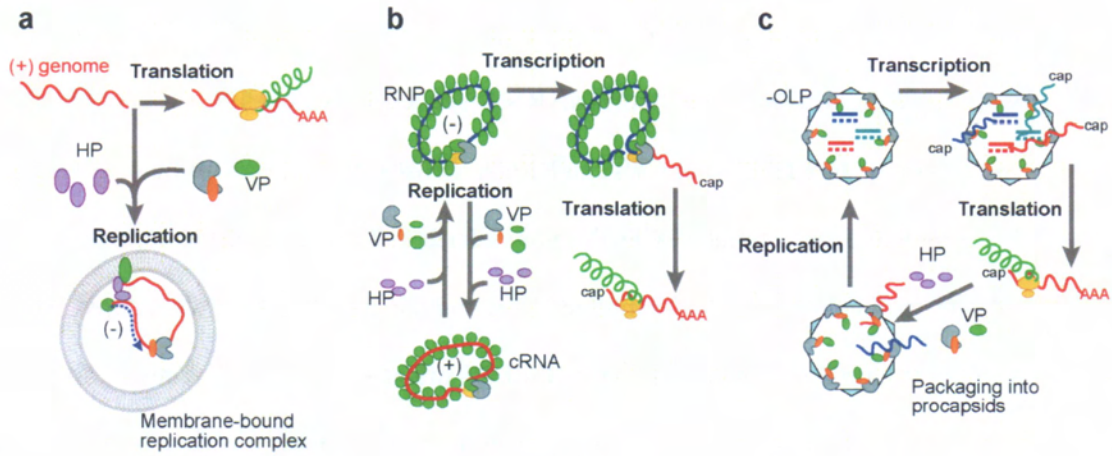


Fig 1.3.Schematic representation of viral replication.

Simplified replication scheme for (A) positive, (B) negative and (C) dsRNA viruses. (A) The (+) strand genome is infectious and serves as template. Replication occurs within membrane vesicles in a multiprotein complex. The (+) strand RNA is also translated by host machinery proteins into viral proteins, which then play a role in viral replication. (B) The (-) strand RNA is transcribed as the first step of replication. The viral genome is transferred into the infected cell in the form of a ribonucleoprotein complex (RNP). RNA replication proceeds via a (+) strand intermediate. The (-)

strand RNA genome requires RNP that is associated with viral proteins. The viral proteins cause the RNP to adopt the helical or circular conformation. (C) The infectious rotavirus particle make up of three layers of protein that contains a genome of segmented dsRNA. Upon infection, the rotavirus virions loose its outer to form a double layer particle (DLP) that serves as transcriptase for genome transcription. The capped mRNAs are released into the cytoplasm through channels at the vertices of the DLPs, which could serve as template for the synthesis of new dsRNA. Host proteins (HP), viral proteins (VP), ribonucleoprotein (RNP) and virus particles without outer layer (OLP). Adapted from Ortin et al., 2006.

1.2.4 Host Factors in RNA Replication

There is mounting evidence to support the interaction between host cellular proteins and viral RdRp or with the replication complex. Some of these identified host proteins may modulate the protein-RNA or RNA-RNA interactions. For example, host proteins that were identified from poliovirus and Qb infected cells can alter the secondary structure of RNA to promote RNA binding to replicase (van Dijk et al., 2004). A ribosomal component EF1-α was identified to interact with 3Dpol (poliovirus) and 5' of viral RNA for efficient RNA-RdRp binding (Harris et al., 1994) whereas a subunit of the eukaryotic initiation factor 3 (EIF-3) was identified to modulate the activity of brome mosaic virus (BMV) RdRp (Quadt et al., 1993). In Flaviviridae, two cellular proteins have been identified that play a role in HCV viral replication, these proteins are (1) geranylgeranylated FBL2 protein that binds to NS5A, a virus encoded phosphoprotein (Wang et al., 2005) and (2) cyclophilin which is a peptidyl-prolyl-cis-trans isomerase that binds to the replicase, NS5B (Watashi et al., 2005). Inhibitions of these two proteins have been shown to be detrimental to viral replication (Wang et al., 2005; Watashi et al., 2005). Other examples include the human nucleolar protein, nucleolin, that was reported to bind HCV NS5 RdRp and able to transport HCV NS5 RdRp into the

nucleus and inhibits its catalytic activity (Shimakami et al., 2006). In DENV, the human La autoantigen was found to interact with NS3, NS5, 5' and 3'UTR of the positive strand RNA and the 3'UTR of the negative strand RNA (Garcia-Montalvo et al., 2004).

1.3 Viral Replication induces membrane rearrangement

For many (+) RNA viruses including flaviviruses replication occurs in association with 50-70nm diameter membranous vesicles or spherules (Schwartz et al., 2004). The spherules are formed in the lumen of specific secretory compartments, endosomes, mitochondria endoplasmic reticulum (ER) and other organelles which contain the replication complex. The site for spherule formation and membrane rearrangements varies in different (+) RNA viruses (Ahlquist et al., 2003; Westaway et al., 1997b). Flavivirus replication factors and double stranded RNA replication intermediates colocalize in a vesicle packet of 50-100 nm that are enclosed in a second membrane layer (Westaway et al., 1997b) whilst coronaviruses replication is associated with clusters of 150-300 nm of double layered membrane vesicles (Egger et al., 1996).

Membrane rearrangements associated with (+) strand RNA virus replication has been shown in the BMV. BMV could encode 2 viral replication factor proteins that direct viral RNA replication. Viral replication factor 1a (pro 1a) contains a helicase and MTase enzymatic activity whereas viral replication factor 2a (pro 2a) contains a central polymerase domain (Chen and Ahlquist, 2000). Pro 1a localizes to outher perinuclear ER membranes where it induces membrane lipid synthesis and vesicular invagination that serves as compartment for RNA replication (Chen and Ahlquist, 2000). Subsequently, pro 1a recruits viral RNA templates and pro 2a to the compartments. The authors show that the balance of the BMV replication factors 1a and pro 2a could modulate the structure of membrane rearrangements.

The presence of high levels of pro 2a protein could shift the membrane rearrangements associated with RNA replication from small invaginated spherules to large, karmellae-like and multilayer stacks of appressed double membranes that support RNA replication as efficiently as spherules (Schwartz et al., 2004). In contrast, spherules were induced by the presence of pro 1a only. Interestingly, Flavivirus RNA replication also localizes to packet of 50-100 nm vesicles surrounded by a second membrane, including DENV (Westaway et al., 1997b). These flaviviruses vesicle packets appear similar to spherules invaginated into ER or mitochondrial lumens induced by BMV infection as observed in an EM study. Therefore, the double membrane rearrangement of ER in BMV that is associated with RNA replication may represent related structures formed by similar protein-protein and protein-membrane interactions in other (+) strand RNA viruses replication (Schwartz et al., 2004).

Recently, the same group of authors also revealed a 3 dimensional model of mini organelle induced by flock house virus (FHV) replication that was reconstructed from EM tomography (EMT) (Kopek et al., 2007) (Fig 1.4). The FHV contains only the FHV RNA replication factor (protein A) and replication occurs at mitochondrial membranes (Fig 1.4). The results show that the newly synthesized FHV RNA and single FHV RNA replication factor localized in the specific membrane vesicles which is inside the outer mitochondrial membrane. A similar observation was also made in flavivirus replication (Kopek et al., 2007). The EMT image shows that each vesicle is connected to the cytoplasm by a single necked channel that able to import ribonucleotide substrates and export RNA product. With these observations, the authors proposed that these vesicles function as replication compartments which protect RNA from competing activate cycle and host defenses.

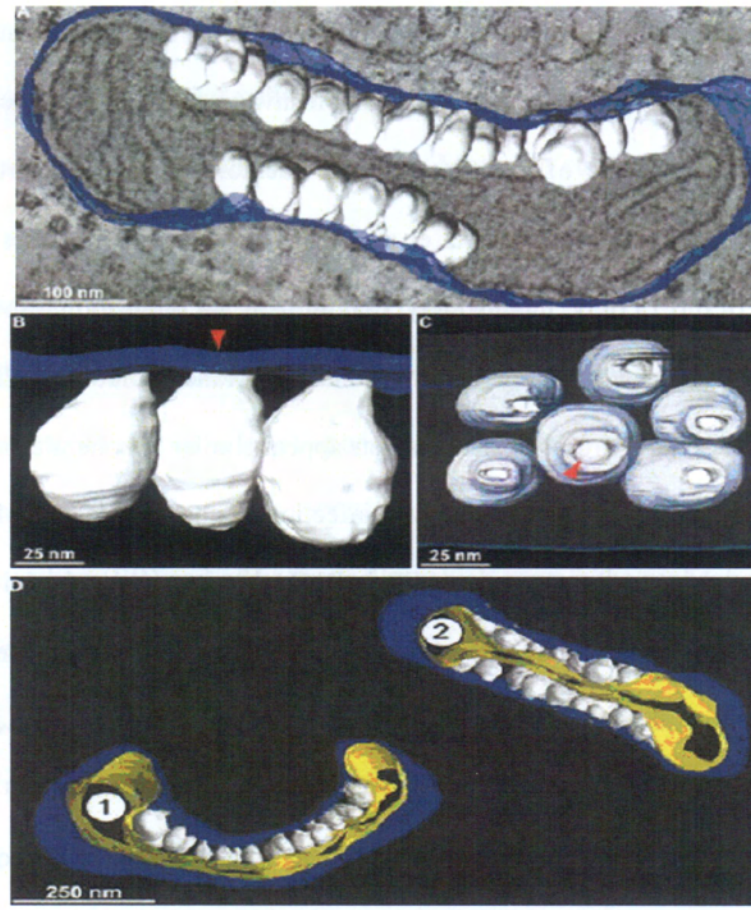


Fig 1.4.3D maps of FHV induced mitochondrial spherules.

3D Maps of FHV induced mitochondrial spherules generated based on slice of EMT reconstructions. For details on the figure refer to Kopek et al., 2007. Blue represents outer mitochondrial membrane, white represents FHV spherules and yellow represents the inner-mitochondrial membrane. (A) Image merged from the outer membrane and spherules of the mitochondrion. (B) A close-up view of the connections between the outer mitochondrial membrane and the spherules. (C) A view of 90° rotation of (B) that shows the channels connect ing the spherule interior to the cytoplasm. (D) 3D maps of the FHV induced mitochondrial spherules that show the inner membrane. Adapted from Kopek et al., 2007.

1.4 Molecular biology of dengue virus

Flaviviruses are small enveloped positive-strand RNA viruses with a diameter of ~ 50-100 nm (Fig 1.5). The virus particle consists of a poorly ordered cage like nucleocapsid. This nucleocapsid is surrounded by a 40 Å lipid bilayer and is anchored with the membrane (M) and envelope (E) proteins (Fig 1.5). The viral RNA genome is housed within a cage that comprises multiple copies of the capsid protein (C).

The RNA genome of flavivirus is approximately ~11 kb in length and contains an open reading frame that encodes a single polyprotein precursor (Fig 1.6) (Coia et al., 1988). The 5' and 3' untranslated regions (UTR) are ~ 90-130 and 430-760 nucleotides (nt) long, respectively. Both of these 5' and 3' UTRs contain RNA elements that act like promoters that are essential for the efficient translation and replication of the flavivirus genome (You and Padmanabhan, 1999; Chen et al., 1997). At the 5' end, flavivirus RNA contains a type 1 cap structure ($m^7G'pppA^mp$) and the 3' end terminates with a CU_{OH}. This is in contrast to the hepaci and pestiviruses which contain polyadenylated tail at the 3'end and internal ribosomal entry site (IRES) at the 5' end, respectively (Lindenbach and Rice, 2003).

Translation of the flavivirus genome results in the production of a polyprotein precursor (~3400 amino acids) (Fig 1.6). The polyprotein is co- and post-translationally processed by viral and host cellular proteases to yield 10 mature viral proteins: three structural proteins comprising of capsid (C), premembrane (prM)/membrane (M), envelope (E) and seven nonstructural proteins, NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5 (Table 1.1). The C, M and E proteins makeup the virus particle with M and E proteins forming the virus coat and the capsid protein, involved with packaging of the viral genome and forming the nucleocapsid core (NC). The NS proteins are assumed to be involved primarily in the

replication of viral RNA as a part of the replication complex. Other functions of the NS proteins include expression, assembly, and release of viral particles and also potentially function in evading the immune response to viral infection (Kummerer and Rice, 2002; Lindenbach and Rice, 1999; Liu et al., 2003; Liu et al., 2005).

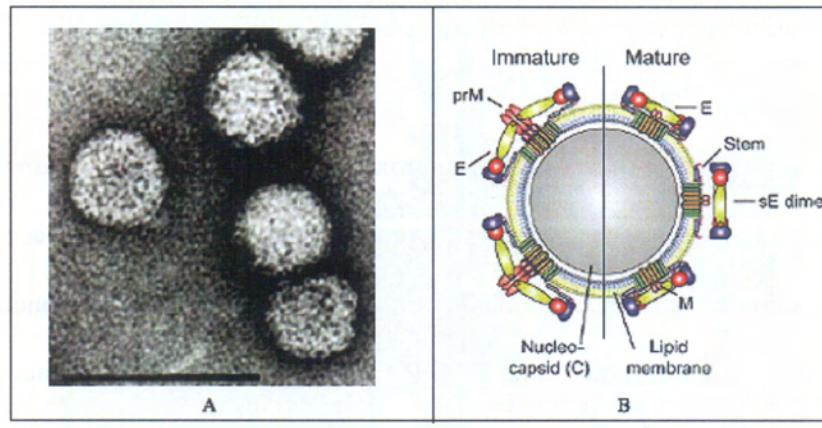


Fig 1.5.Dengue virus particle

(A). Electron micrograph of dengue virus particle, bar represents 100nm. Adapted from (Lindenbach and Rice, 2003). (B) Schematic representations of dengue virus particle that shows the immature or mature form. (Stiasny and Heinz, 2006).

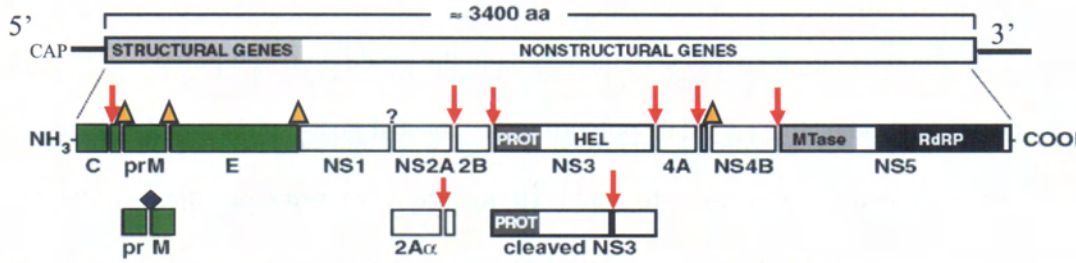


Fig 1.6. Flavivirus genome organization and polyprotein processing.

Viral NS2B-NS3pro, signal peptidase and furin cleavage sites are represented by arrow, triangle and diamond respectively. The NS1-2A processing by unknown ER resident enzyme is represented by question mark.

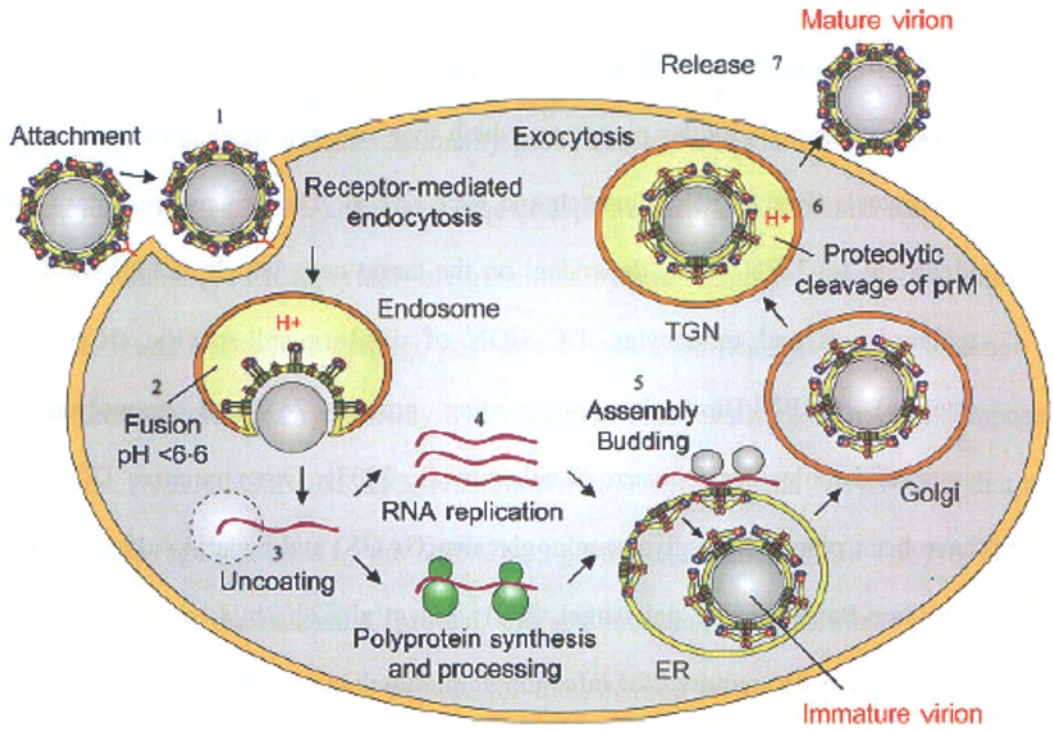


Fig 1.7. Flavivirus replicative cycle.

(1) Virus enters the target cell by receptor mediated endocytosis. (2) E proteins experience structural change induced by the acidic pH in the endosome and lead to membrane fusion. (3) Uncoating. (4) Translation of plus stranded RNA genome to initiate replication, with co- and post translational proteolytic processing of viral polyprotein by viral and host proteases. (5) Virus assembly in the ER leads to the formation of immature virion that is transported to Golgi. (6) Proteolytic cleavage of prM that leads to the formation of mature virion. (7) Mature virion released by exocytosis. Adapted from (Stiasny and Heinz, 2006)

1.4.1 Flavivirus Replication Cycle

1.4.1.1 Entry and uncoating

Flavivirus particle enters host cells via the interaction between the viral E/prM of surface glycoproteins and cellular receptors, which then triggers a process called receptor mediated endocytosis (Fig 1.7) (Lindenbach and Rice, 2003). The primary receptors that have been identified for DENV vary dependent on the target cells which include the Fcγ receptors of macrophages and monocytes, DC-SIGN of dendritic-cell-specific (ICAM-grabbing non-integrin), GRP78/Bip (glucose-regulating protein 78) and lipopolysaccharide CD14 associated molecules (Navarro-Sanchez et al., 2003). Other putative DENV receptors that have been reported are glycosaminoglycans (GAGS) and heparin sulfate (Hung et al., 1999; Martinez-Barragan and del Angel, 2001; Lin et al., 2002b; Lee and Lobigs, 2000). These receptors are also required for infection in mammalian cells.

Following endocytosis, the RNA genome is released into cytoplasm of the host cell through fusion of the viral envelope with the host cell endosomal membrane. This process is mediated by the E protein and its fusion peptide (Allison et al., 1995; Rey et al., 1995). First, the homodimeric E protein binds to a receptor and is internalized in the endosome. Second, the conformational change causes the virus in an open form and thereby exposes the fusion peptide, bringing the endosomal and viral membranes into close proximity and enables the formation of homotrimeric E protein. This step allows the insertion of virus into host cell membrane and subsequently is followed by the change of E protein into monomers (Modis et al., 2003). The virus particle is then disassembled by explusion of nucleocapsid into the cytoplasm.

1.4.1.2 Translation and processing

Once the positive single stranded RNA is uncoated and released into cytoplasm, the 5' UTR directs the RNA to the ribosomes, where the viral RNA is translated by into single polyprotein that is co- and post-translationally processed by viral and host proteases into at least 10 proteins (Fig 1.6 & 1.7). The cleavage at C/pr-M, prM/E, E/NS1 and the C terminal of NS4A in the lumen of ER reportedly are generated by host signal peptidase. This cleavage releases prM, E and NS1 which undergo post-translational modifications in the ER lumen as they move through the secretory pathway. The majority of the C terminal is mainly composed of NS proteins (NS2A/NS2B, NS2B/3, NS3/NS4A, NS4A/NS4B, and NS4B/NS5) that are generated by cleavage mediated by the NS2A/NS3 protease (Brinton, 2002). The NS2B/NS3 protease also cleaves the nascent C protein, thereby releasing its C terminal hydrophobic domain which functions as a signal peptide for translocation of prM into the ER lumen to produce mature C (Amberg et al., 1994).

1.4.1.3 RNA replication

In the replication complex model, it is assumed that the consensus components NS3, NS2A, NS4A, and NS1 assemble with NS5 near the conserved 3'terminal stem loop when translation of the (+)RNA in cis is completed (Westaway et al., 2003). This is followed by the circularization of the complementary sequence (CS) at the 5' and 3' terminal stem loops and the NS proteins that make up the replication complex assembled at the 3' terminal. Viral replication is catalyzed within a replication complex that comprises (1) NS5, the RNA dependent RNA polymerase and methyl transferase (MTase), (2) NS3, the viral RNA helicase and protease that is needed for unwinding the replicative form, RF (dsRNA) and also for processing viral polyprotein and (3) other host factors (Bartholomeusz and Wright, 1993; Mackenzie et al., 1998; Nova-Ocampo et al., 2002). The replication complex is believed to be

tethered to the membrane by NS4A, an integral membrane protein. The viral replication occurs within the induced perinuclear vesicles called vesicle packets (VP) that reside in Tran Golgi network (TGN).

The first step is that a negative RNA strand is synthesized from the positive RNA genome to form the double stranded RNA, known as the Replication Form (RF) (Chu and Westaway, 1985; Chu and Westaway, 1987). The RF is the template for asymmetric and semi-conservative RNA synthesis which results in the formation of the replicative intermediate (RI) with the nascent single stranded RNA tails. Upon completion of strand synthesis, the tail of RI is resolved to generate single stranded (+) RNA and a RF. The positive strand synthesis is reportedly 10 times more efficient than that of negative strands (Chu and Westaway, 1985). Recently Uchil et al., 2006 described the architecture of the flaviviral replication complex (Uchil et al., 2006). They have shown that a double membrane compartment encloses the RF of DENV, WNV and JEV. The authors also demonstrated that the newly synthesized viral genomic RNA associated VP are oriented outwards with the RF located inside the vesicles.

1.4.1.4 Assembly and Release

After replication, the viral genome is encapsulated by C and transferred to the ER (Fig 1.7). The nucleocapsid is processed in ER and enveloped with a lipid bilayer. The E and prM forms heterodimer that binds nucleocapsid and promotes the viral assembly. In addition, role of the NS proteins in viral assembly have been reported. Mutation in NS2A of YFV was shown detrimental to viral assembly (Kummerer and Rice, 2002). Similarly, NS3 is proposed to link RNA and structural proteins during assembly (Liu et al., 2002). Viral assembly that occurs in the ER leads to the formation of immature virions and are subsequently transported

to TGN (Fig 1.7). In TGN, the maturation of virion occurs prior to release the viral particles through exocytosis.

1.5. NS5- A Multifunctional Protein

NS5, the main focus in this work, is the largest of the DENV proteins with a molecular mass of 104 kDa (Fig 1.8). Sharing a minimum of 70% amino acid sequence identity across the four DENV serotypes, NS5 is the most conserved viral protein. It has many critical functions including replication, capping of the RNA, and possibly host cell gene regulation. Based on structural and biochemical studies, three functional domains have been identified in NS5 (Fig 1.8). The N-terminal S-adenosyl-methionine methyltransferase (SAM) domain of several flavivirus, spanning ~ 300 amino acids (Koonin, 1993), has been expressed as a soluble active RNA capping enzyme (Fig 1.8) (Egloff et al., 2002). RNA capping is a process that involves RNA triphosphatase (hydrolyze 5' RNA to a diphosphate end; NS3 has been suggested to carry RTPase activity) (Lindenbach and Rice, 2003), an unidentified guanylyltransferase (cap RNA with GMP), RNA guanine-methyltransferase (methylate at the N-7 position of guanine) and 2'-O methyltransferase, 2'-O MTase (methylate at the ribose 2'-OH). A central S-adenosyl-methionine-dependent methyltransferase (MTase) core structure sits within a "cradle", shaped by N-terminal and C-terminal sub-domains (Egloff et al., 2002). Recent studies demonstrate that the active site of MTase of several flavivirus (DENV and WNV) is able to conduct both guanine N-7 and ribose 2'-O methylations with different mechanism, despite the fact that MTase crystal structures (WNV and DENV-2) only displayed a single substrate binding site for S-adenosyl-L-methionine (Adomet) as the methyl donor (Fig. 1.9) (Egloff et al., 2002; Ray et al., 2006). These two methylation events are required for proper formation of the cap structure, which is recognized by the host cell translational machinery.

The C-terminal region of NS5 contains five amino-acid sequence motifs that form the signature of RNA-dependent RNA polymerases (RdRp) (Fig 1.8). Working with other viral proteins and as yet unidentified host cell proteins, the NS5 polymerase domain is responsible for synthesizing a transient double stranded replicative RNA intermediate. The newly synthesized minus strand serves in turn as a template, allowing the RdRp to synthesize additional plus strand genomic RNA (Kapoor et al., 1995; Nomaguchi et al., 2003; You et al., 2001a). Several activity assays for the corresponding Flaviviridae RdRp domains have been described. For example, Kunjin virus (KV), Hepatitis C viruses (HCV) and bovine viral diarrhea virus (BVDV) (Khromykh et al., 1998; Steffens et al., 1999; Tan et al., 1996).

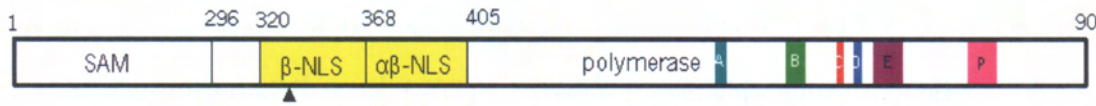


Fig 1.8.Schematic representation of the distribution of the functional domains in the DENV NS5 protein.

The FL NS5 protein has three major functional domains. The N-terminal MTase spans residues 1 to 296. The NLS has been divided into a/bNLS or b-NLS. The b-NLS region is thought to interact with the NS3 viral helicase. The six conserved amino acid sequence motifs within the RdRp domain (residues 273-900) are colored respectively and P denotes the priming loop site.

Fig 1.9.Crystal structure of DENV NS5 MTase

The structure is in complex with S-adenosyl-L-homocysteine (AdoHcy) shown in ribbon and ball and sticks representation respectively. The AdoHcy is the product of methylation of reaction, which AdoMet is the substrate. Adapted from (Egloff et al., 2002).

1.5.1 NS5 Nuclear Localization: its implication in viral replication

NS5 nuclear localization has its implication on viral replication that is also associated with NS5 phosphorylation. Reed et al., 1998 revealed that NS5 of HCV, YFV and BVDV can be phosphorylated by serine/threonine kinases (Reed et al., 1998). This suggested that the same or similar host kinase could phosphorylate these viral NS5s during viral replication. Kapoor and co-workers demonstrated that interaction between DENV NS3 and NS5 proteins during viral replication is modulated by NS5 nuclear localization and phosphorylation (Kapoor et al., 1995). They reported that NS5 can be phosphorylated on serine residues. The phosphorylated NS5 protein is dissociated from NS3 with NS3 undergoing autoproteolysis and then NS5 is transported into nucleus by elaborated host machinery. In contrast, the non-phosphorylated NS5 interacts with NS3 within the replication complex where viral RNA replication occurs. In addition, NS5 was shown to be localized into nucleus during the late cycle of infection and its role remains unknown.

Forwood et al., 1999 were the first to demonstrate that a 37-amino acid interdomain (residues 369-405) of DENV 2 NS5, which lies between the MTase and RdRp domains contains a nuclear localization signal (NLS). This NLS region is able to target β-galactosidase (a large cytoplasmic protein that is usually excluded from the nucleus by its size) to the nucleus, indicating that the same site contains signal responsible for NS5 nuclear localization (Fig 1.10) (Forwood et al., 1999). A putative protein kinase CK2 phosphorylation site within the NLS region was also discovered. Furthermore a CK2-phosphorylated NS5 was found to inhibit NS5 nuclear localization (Fig 1.10), suggesting that the nuclear transportation is also modulated by a cytoplasmic retention mechanism (Brooks et al., 2002). Two nuclear localization sequences (NLS) between amino acid residues 320-405 (b-NLS: 320-368 and ab-NLS: 369-405) of NS5 was mapped out using Y2H system. The b-NLS and ab-NLS were reported to interact with importin β and importin αβ, respectively (Brooks et al., 2002) (Fig 1.10). Mutational studies on ab-NLS revealed that K387/K388/K389 changes to alanine impaired NS5 nuclear accumulation and virus production. This implies that NS5 nuclear localization via phosphorylation play a critical role in viral replication (Pryor et al., 2007; Hanley et al., 2002; Brooks et al., 2002). In addition, the NLS signals were shown to be able to recognize cellular factors for nuclear transport. The precise role of the nuclear translocation event and its role in dengue pathogenesis is still not clear.

Interestingly, a segment of 20 amino-acids that is strictly conserved in all flaviviruses also lies between amino acid residues 320-368 of NS5. This region of NS5, which is involved in binding importin β, is also thought to interact with NS3 (Fig 1.10). The association between NS5 and NS3 is expected to modulate their respective enzymatic activities. Indeed, NS5 was reported to stimulate the nucleotide triphosphatase (NTPase) and RNA triphosphatase (RTPase) activities of NS3 (Yon et al., 2005). Recently, Uchil et al., 2006 presented the

hypothesis that a host nucleus could functions as a site for functionally active flaviviral replicase complex. About 20% of the total RdRp activity from flavivirus infected cells (JEV, WNV and DENV) was found to reside within the nucleus, which is consistent with nuclear co-localization of replicase NS3-NS5 complex. On the other hand, HCV RdRp has been demonstrated to interact with a protein kinase PRK2 *in vivo* (Kim et al., 2004). The kinase PRK2 was found to phosphorylate RdRp's fingers domain and accompanied with the increase of viral replication. Also phosphorylation on Turnip Yellow Mosaic virus (TYMV) RdDp (RNA dependent DNA polymerase) was found both *in vivo* in viral infected cells and in the *in vitro* baculovirus expressed protein (Jakubiec et al., 2006). One of the phoshorylation sites was mapped out within the palm subdomain of this RdRp, suggesting a possible role of phosphorylation in (+) strand RNA virus replication.

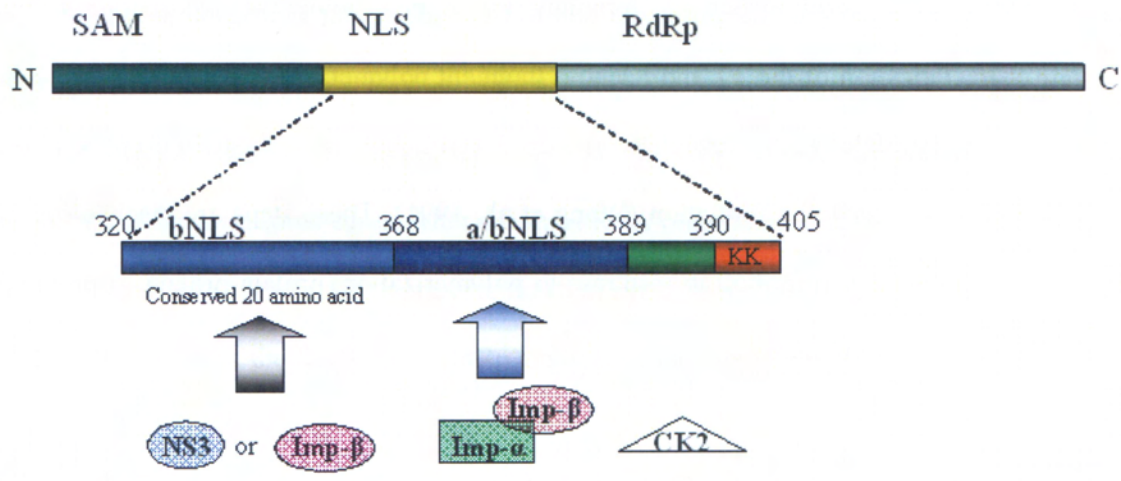


Fig 1.10.Schematic representation of the domain structure of NS5 shows the NLS region and their partner protein binding sites.

The 20 conserved amino acids in the bNLS site are potentially interact with importin β or NS3 viral helicase. The abNLS putatively interacts with importin $\alpha\beta$ and protein kinase CK2. Adapted from (Pryor et al., 2007).

1.6 An Unified Polymerase Mechanism

Despite the structural variations across the four polymerase families (see section 1.8), all polymerases share a common polymerization mechanism (Fig 1.11). The polymerase enzyme (E) binds template/primer (p/t) to form the E-p/t complex (step 1). The substrate then binds the E-p/t complex to form E-p/t-S complex (step 2), which is a rate limiting step, before convertion to an activated complex E'-p/t-S (step 3). The rate limiting step is believed to be caused by the conformational change that polymerase undergoes, but this may not true for all viral RdRps. After this, phosphodiester bond between the growing chain and the incoming nucleotide substrate occurs via the nucleophilic attack by the 3'-OH group of the primer terminus on the α-phosphate of incoming substrate (step 4). The polymerase is thought to undergo a second conformational change that leads to the release of the leaving group pyrophosphate (PPi). The enzyme can then dissociate from the E-p/t or translocate the substrate to form a new 3' terminus for a new round of incorporation. Functional characterization of the T7 and Klenow fragment polymerases revealed that step 3 and 4 are crucial for fidelity, by reduction in step 3 (formation of activated complex E'-p/t-S) and phosphodiester bond formation (Wong et al., 1991). These steps are mechanistic processes that allow protein motion in response to polymerization, protein-protein or protein/RNA or protein DNA interactions.

1.6.1 Viral RNA-dependent RNA polymerization

RNA synthesis initiation can be either through de novo or primer dependent process (Fig 1.12). Nucleotides at the 3' end of template are denoted as T1, T2,..., in the 3'-5' direction. The newly synthesized nucleotides are denoted as D1, D2... in the 5'-3' direction, where D1 base pairs with T1 and so on. The D1 nucleotide is also referred to as the initiation nucleotide (NTPi) and D2 is referred as NTPi +1. Biochemical analysis revealed that GTP is preferred

as the D1 nucleotide for RdRps of flaviviruses even though adenine is the first nucleotide of both (+) and (-) strand of flavivirus RNA (Selisko et al., 2006).

1.6.1.1 De novo initiation

Viruses that belong to (+), (-) or double stranded RNA classes can initiate de novo replication. De novo initiation requires the following components for replication (1) the RdRp, (2) the RNA template, (3) the initiation nucleotide (D1 or NTPi, usually either a GTP or ATP) and (4) the second nucleotide (D2 or NTPi +1)(Dreher, 1999; Ball, 1995). The first phosphodiester bond is formed between D1 and D2, of which the D1 provides the 3' hydroxyl for the addition of new nucleotide. Kinetic analysis revealed that HCV RdRp has higher K_m for the NTPi than for the following of that NTP +1. The initiation machine is thought as a stable complex that is needed to regulate accurate positioning of the nucleotide. In the de novo initiation, no genetic information is lost and no additional enzymes from host or virus are required to synthesize or cleave a primer during the replication (van Dijk et al., 2004). Elongation is the immediate next step after the de novo initiation. However, abortive initiation can occur in some cases. After the short RNA oligonucleotides is synthesized, it can be used to serves as primer for the next replication cycle or replace the D1 *in vitro* (van Dijk et al., 2004). Most of the flaviviruses RdRp are believed to conduct de novo initiation *in vivo*.

1.6.1.2 Primer dependent initiation

Some viruses use a primer for initiation of polymerization. The primer can be classified as (1) protein primers where the hydroxyl group of the amino acid forms a phosphodiester bond with the D1, as seen in the Foot Mouth Disease Virus (FMDV) RdRp template-primer complex structure (Ferrer-Orta et al., 2006a), (2) oligonucleotides that are cleaved from the 5' end of a capped cellular mRNA, as used by many negative strand RNA viruses (Hagen et al.,

1995), (3) a short RNA primer that is derived from abortive initiation process (Mcclure, 1985) and (4) the 3' end of the template folds back to be a primer, also known as copy back mechanism (Fig 1.12) (Laurila et al., 2002). However, back priming that is also observed in HCV, BVDV and DENV RdRp and this could be an artifact of *in vitro* system.

Recently, the structure of FMDV RdRp-protein primer (vPg) has been determined, as an example of protein-primers dependent enzyme (Ferrer-Orta et al., 2006a). The protein primer (vPg) is linked to the 5' end of the viral genome and viral RdRp can catalyzes the nucleotidylation of vPg in the presence of RNA template. In addition, RdRps of poliovirus or calicivirus are those that use oligonucleotide primer dependent initiation. This structural feature that enables the calicivirus and poliovirus RdRps to carry out the primer dependent initiation without the use of β-hairpin adjacent to the active site. In contrast, a β-hairpin is observed in the case of HCV (or priming loop in DENV and WNV). This β-hairpin is thought to occupy the active site and occlude the primer binding space.

1.6.1.3 Dengue RdRp Initiation

Several studies have focused on the detailed characterization of DENV RdRp. Replicase isolated from the viral infected cell culture has been shown to be active on the subgenomic particles (770 nucleotide RNA template), which contains both the 5' and 3' UTR (You and Padmanabhan, 1999). A product (1X) that is the same size as the template was produced, through the *de novo* initiation mechanism. A product (2X) the size of the template was also shown. This product is a hairpin RNA that contains twice the size of template, presumably derived by copy back mechanism. The results suggest that the viral replicase could perform both de novo and primer dependent initiation *in vitro*. Interestingly, purified FL NS5 also resulted in a similar observation, without other replication complex components such as NS3.

Recently, Selisko et al., 2006 also showed that DENV RdRp is active both on specific heteropolymeric and homopolymeric templates and MTase domain does not interfere with the polymerase activity (RdRp v.s FL DENV) (Selisko et al., 2006). In kinetic assays, the 5' and 3' terminal regions which comprise the 5' UTR, 3' stem-loop structure, 5' and 3' cyclization motif were shown to be essential in the long range interaction for viral replication. These elements are thought to be important for minus-strand RNA synthesis modulation, which is also observed in Kunjin virus (KUNV). In addition, mutational analysis further shows that mutated dinucleotide CU (U to C) at 3'end reduces the activity drastically whereas mutation on the penultimate (C to U) did not affect the template activity (Nomaguchi et al., 2003). A model was proposed by the authors where the active site of RdRp switches from a "closed" conformation during de novo initiation to an open conformation for elongation of a double stranded template primer.

In a separate study, lower temperatures were found to promote de novo RNA replication and higher temperatures appear to favor the copy back products (Ackermann and Padmanabhan, 2001). A high concentration of GTP is also required for the formation of RdRp de novo initiation complex, regardless of the template sequence within the first 6 nucleotides from the 3' end. For all flaviviruses, GTP is the second incorporated nucleotide during positive or negative RNA synthesis. Two hypothesis for the potential role of GTP role were discussed (1) GTP is actually the initiating nucleotide for negative strand synthesis or (2) GTP has putative allosteric binding site to allow the enzyme to adopt closed conformation for de novo initiation (Selisko et al., 2006).

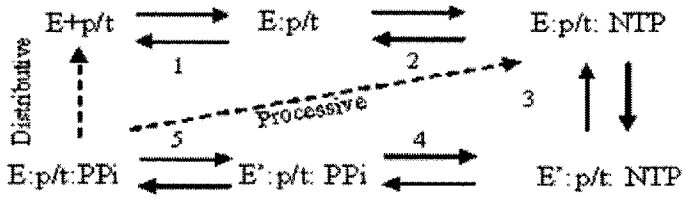


Fig 1.11.Kinetic model for nucleotide incorporation.

A model shared by all polymerase. See text for details. Enzyme (E), template/primer (p/t), substrate (NTP), pyrophosphate (ppi). The formation of activated complex E'-p/t-S (step 3) and phosphodiester bond formation (step 4) is believed to be the rate limiting step due to polymerase conformational change.

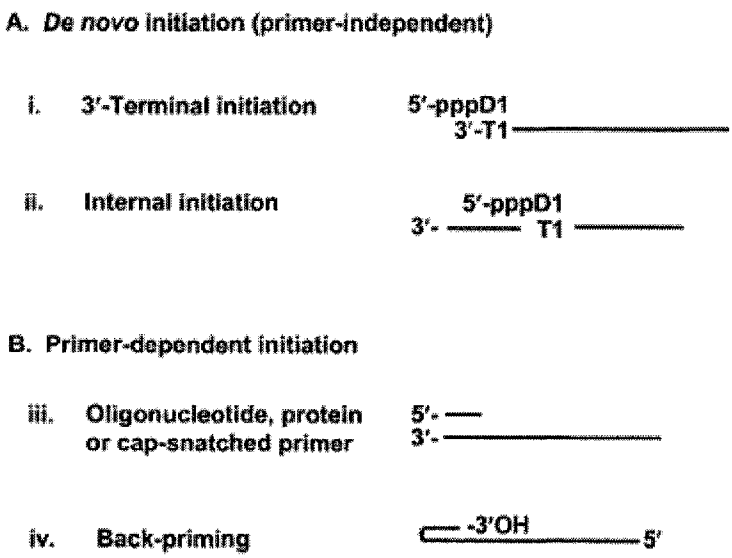


Fig 1.12.Schematic view of initiation mechanisms of viral RdRp.

(A) De novo initiation. (i) From the 3'termnial of the viral genome (ii) internal to the template during subgenomic synthesis. (B) Primer dependent initiation. (iii) Primer can be a oligonucleotide, protein primer or a capped primer. (iv) Copy mechanism or back-priming with the 3' terminal folding back. Adapted from van Dijk et al., 2004.

1.7 Role of 5' and 3' UTR in viral replication

5' unstranslated region (UTR) is ~100 nucleotide long, which is located at the 5' terminal RNA. The 5'UTR is conserved in the RNA of all four DENV serotypes but not in the flaviviruses (Brinton and Dispoto, 1988). 5' UTR has a cap structure at the 5' end. Deletion studies on this region of DENV4 show the derivative impairs viral replication (Cahour et al., 1995). The 3' UTR is longer than 5'UTR and contains ~450 nucleotides. The 3' UTR lack a poly (A) tail and ends in a very conserved 3' stem-loop (3'SL) in many flaviviruses. This 3'SL is thought as a common characteristic that is required for viral replication (Brinton et al., 1986; Zeng et al., 1998). The WNV 3' UTR has been shown to interact with translation elongation factor, EF-1α, but its biological role remains unclear (Chen et al., 1997). Furthermore, 3' UTR of JEV viral RNA has also been shown to interact with NS3 and NS5 (Chen et al., 1997). Upstream of the 3'SL, there is an essential RNA element for viral replication-conserved sequence (Men et al., 1996). This element contains a complementary sequence CS (3' CS) that is complementary to a sequence at the 5' end (5' CS) and present in all mosquito borne flavivirus genomes (Hahn et al., 1987). RNA-RNA interactions in the 5'-3' long range have been shown to be necessary for RNA replication of WNV, KUNV, DV and YFV (Hahn et al., 1987; Khromykh et al., 2001a; Corver et al., 2003; Alvarez et al., 2005). The base pairing of 5'-3' CS of KUNV was shown to be feasible from the computational thermodynamic analysis (Khromykh et al., 2003). This is further supported by effective RNA synthesis only in the presence of both 5' and 3' CS in DENV replication (You et al., 2001b).

Mutational analysis on 5' or 3' CS has been conducted using KUNV virus replicon. The mutant resulted in the abolishment of RNA amplification, due to loss of complementarity between 5' and 3' CS (Khromykh et al., 2001a). Reconstitution of the base pairing could restore viral replication, suggesting a functional role for 5'-3' end interaction in the viral

genome. Recently, Alvarez et al., 2005 also suggested that 5' and 3' CS are essential but not sufficient for RNA interactions, as demonstrated by using atomic force microscopy (AFM) (Fig 1.13) (Alvarez et al., 2005). They identified a second complementary pair sequences (5'-3' UAR) that present in the 5' and 3' UTRs, which serves as a determinant for RNA-RNA interaction (Fig 1.13b) (Alvarez et al., 2005). Filomatori et al., 2006 have also identified a promoter element for RNA synthesis in DENV, which the DENV RdRp can discriminate viral RNA by a specific recognition of a 5' element called Stem Loop A (SLA) (Filomatori et al., 2006). This study found that SLA can be used as a promoter and they proposed the following mechanism for (-) strand RNA synthesis, as depicted in Fig 1.13b (Filomatori et al., 2006). Firstly, the genome circularizes, a step that is mediated by 5'-3' UAR and 5'-3' CS hybridization. Secondly, the viral RdRp binds SLA at the 5'end of the genome and reaches the initiation site at 3' end via long range RNA-RNA interactions to initiate viral RNA replication (Fig 1.13b). In addition, the RNA-RNA interactions between 5' and 3' end of viral genome enhance DENV RNA synthesis only in the presence of intact SLA, which also support the requirement of DENV genome cyclization (Filomatori et al., 2006).

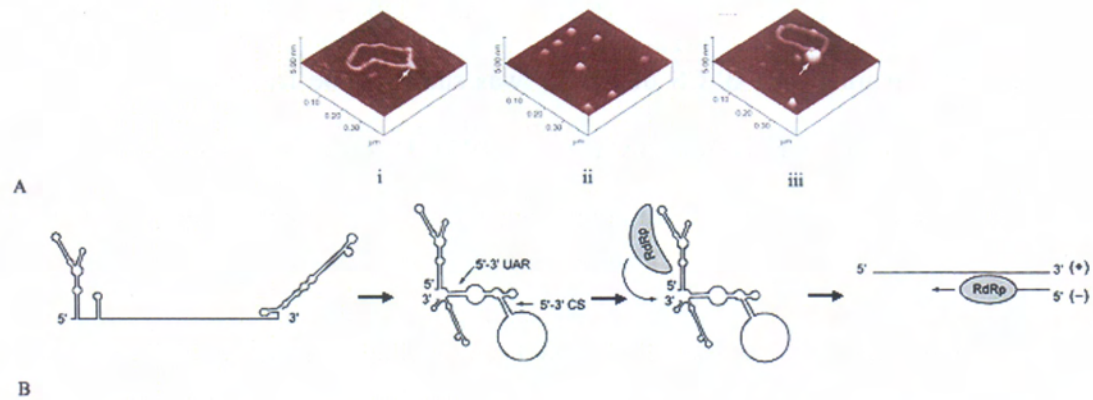


Fig 1.13.Interaction between RNA template and NS5 RdRp

(A) Visualization of RNA-NS5 RdRp interaction by AFM. (i) A representative RNA molecule in a circular conformation, flanked with 5' and 3' ends. (ii) A representative image of NS5 RdRp. (iii) A representative image shows NS5 RdRp bound to RNA. (B) Model for DENV negative strand RNA synthesis. The 5'-3' UAR and 5'-3' CS circularize the viral genome and NS5 RdRp binds to SLA region. Long-range RNA-RNA interaction enables NS5 RdRp be transferred to the 3' end of the genome for replication. Adapted from (Filomatori et al., 2006).

1.8 Structural and Functional Insight into Polymerase

Nucleotide polymerasing enzyme is the key player in catalyzing the formation of polynucleotide sequences. Polymerases use either single stranded or double stranded genome as a template in a processive manner (fast or slow) to produce daughter nucleic acid with high fidelity, depending on the nature of the polymerase. Recently, RNA polymerase (Pol) II, a DNA-dependent RNA polymerase (DdRp), was shown to possess RdRp activity (Lehmann et al., 2007). In this case, Pol II is slower and less processive than DNA-dependent activity. The RdRp activity of Pol II provides a missing link that Pol II evolved from an ancient replicase that duplicated RNA genomes. In a different study, dynamic interaction between helicase and polymerase was also reported to promote processivity in T7 bacteriophage (gp5) (Hamdan et al., 2007), besides its original gp5 processivity on DNA polymerization that increased by thioredoxin (trx) (Tabor et al., 1987). A single copy of these proteins (polymerase + helicase) could process more than 17, 000 nucleotides (Hamdan et al., 2007), compared to an average 700 nucleotides processivity of gp5/trx (Wuite et al., 2000). In addition, *E.coli* DNA polymerase III has a clamp-loading machine for processivity (Indiani and O'Donnell, 2006). The ring-shaped protein forms a sliding clamp that tethers DNA polymerase to the DNA template, thereby enabling the rapid and processive synthesis of both leading and lagging strands at the replication fork. The sliding-clamp protein is loaded on

DNA by a multiprotein clamp-loader complex. In contrast, viral RdRps are error prone during viral replication and hence have a lower fidelity than those of DNA polymerase, due to either the lack of proofreading exonuclease that exist in DNA polymerase.

Polymerases can be sub-classified into four families (O'Reilly and Kao, 1998) (Fig 1.14). The first family is DNA-dependent DNA polymerase (DdDp). A representative is Klenow Fragment (KF) of *E.coli* DNA polymerase I (68 kDa), a monomer that contains a separate structural domain for 3'-5' proofreading exonuclease. The DdDp family can be further sub classified into polymerase α and β (Steitz, 1999). The second family is DNA dependent RNA polymerase (DdRp). This class of protein is represented by T7 RNA polymerase (T7 Pro), a 99 kDa monomeric RNA polymerase from bacteriophage and contains 300 amino acids at the N terminal that folds as a separate domain and is involved in RNA synthesis (Joyce and Steitz, 1995) (Fig 1.14). The third family is RNA dependent DNA polymerase (RdDp), represented by human immunodeficiency reverse transcriptase (HIV RT). This first subunit (66 kDa) of the enzyme is a heterodimer that comprises a polymerase domain and RNase H enzymatic activities on a separate domain. The second subunit (51 kDa) folds into a separate domain and interacts with tRNA which is used as primer for reverse transcription (Joyce and Steitz, 1995) (Fig 1.14). The fourth family is RNA dependent RNA polymerase (RdRp), represented by 3D poliovirus ($3D_{pol}$). This polymerase family can be further differentiated by the mechanism of viral replication initiation (de novo vs. primer dependent) and their respective genus. RdRps are the catalytic component in the viral genome replication machinery. Viral RdRps have "closed hand" conformation, compared to the "open-handed" or U shaped observed in other polymerases (Ferrer-Orta et al., 2006b) (Fig 1.14). The overview of polymerase structure focusess on RdRp (below) because of the three other families of polymerases are not homologous to the RdRps. Nevertheless, the overall shape of

the four families of polymerases resembles that of the right hand with finger, palm and thumb subdomains.

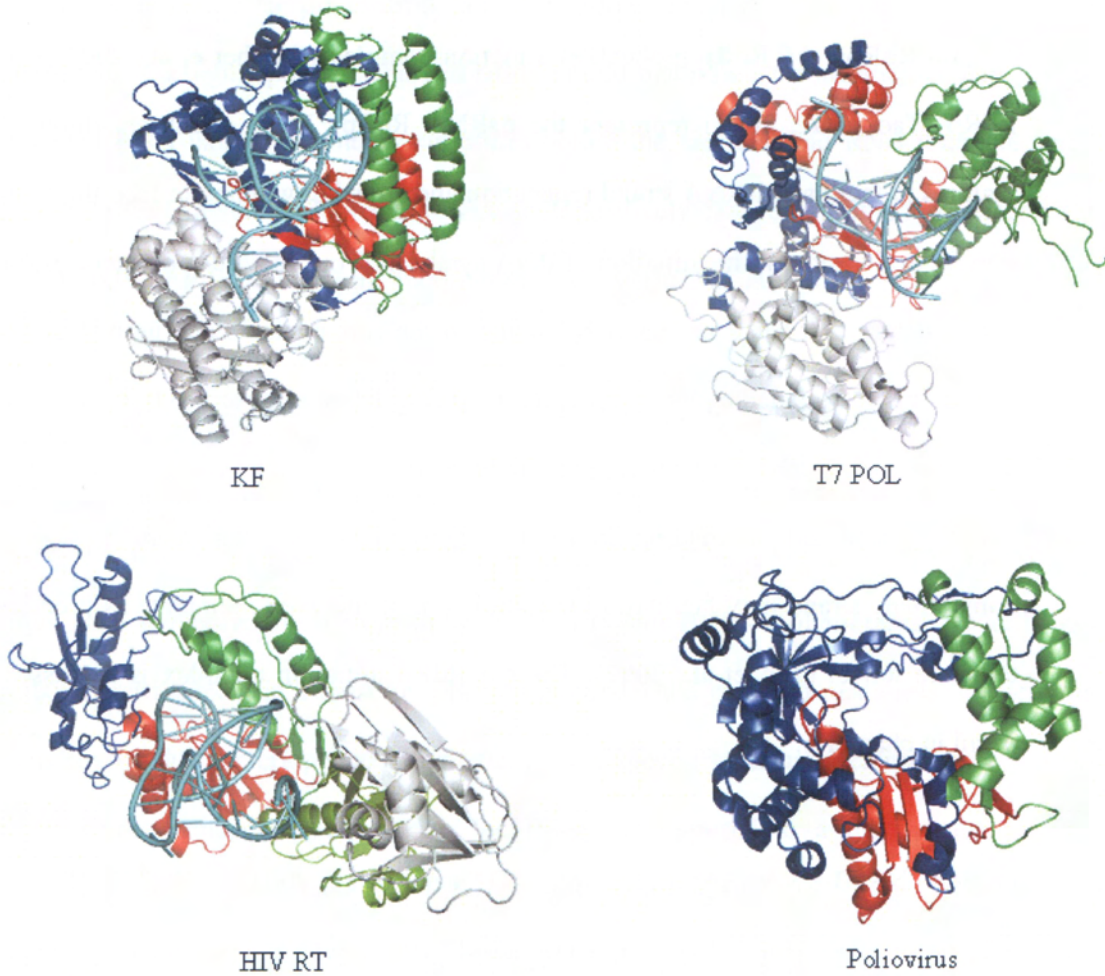


Fig 1.14.Structure of the four polymerase families.

Four polymerase family: KF represents the DdDps family, T7POL represents the DdRps family, HIV RT represents the RdDps family and poliovirus polymerase represents the RdRps family. All the polymerases are represented diagrammatically and colored as- Fingers domain in dark blue, palm domain in red, thumb domain in green and exonuclease domain in gray. In the case of HIV RT, the second subunit p51 is not disclosed for the purpose of clarity. Also the extra 300 amino acids at N terminal for T7POL is not shown. The template- primer for respective polymerases is

represented in light blue. PDB code for respective structures are 1KLN (KF), 1T7P (T7 POL), 1J50 (HIV-RT) and 1TQL (poliovirus).

1.8.1 Template binding- the fingers subdomains

In viral RdRps, Φ6 RdRp replication initiation complex (Butcher et al., 2001) and reovirus RdRp (Tao et al., 2002) represent the dsRNA RdRp. These structures show specialized features to facilitate dsRNA strand replication and strand unwinding to feed the correct strand into the active site for the initiation of RNA synthesis. Four reovirus polymerases form a cage like structure that has four channels leading to the active site. Each individual polymerase core unit is surrounded by its N and C-terminal (Fig 1.15), as observed in the crystal structure. In addition, a 5' RNA cap binding site is located between the entrance and exit tunnels of template. The unique binding location suggests that the attachment of the 5' end of nontemplate strand may facilitates the reinitiation of the 3' end template as it enters into the template tunnel (Tao et al., 2002). The Φ6 RdRp initiation complex is discussed in more detail in chapter 5.

In HCV NS5B-RNA template complex (O'Farrell et al., 2003) or FMDV 3D-Pol-template primer complex (Fig 1.16) (Ferrer-Orta et al., 2004), which represent the ssRNA RdRp, a well defined template entry tunnel in the fingers subdomain towards the active site is disclosed by the structure. In addition, the MTase domain that functions as the 5' capping enzyme at the N terminal of the RdRps distinguishes the flaviviruses RdRps from other polymerases. The MTase and RdRp is synthesized as single polypeptides in the form of full length gene product called NS5 (Egloff et al., 2002; Koonin, 1991). Unlike HIV RT or KF, editing exonuclease is absent in viral RdRps. MTase is also absent in pestivirus (BVDV) or hepacivirus (HCV). In HCV (hepacivirus), NS5 region comprises the NS5A phosphoprotein

that is involved in regulating replication. These HCV proteins were shown to be translated as individual protein respectively. The BVDV RdRp has an N terminal region that fold into a separate domain. This N domain protein is located over the thumb and interacts with the fingertips through a β-hairpin motif (Choi et al., 2006) (Fig 1.17). A potential role has been proposed for the highly positive charged hairpin motif that points toward the template tunnel- to unwind RNA hairpin secondary structures before the single strand RNA template enters the active site (Choi et al., 2006). Likewise such positively charged motif is also observed in HCV, RHDV or RdRps belonging picornavirus genus (Bruenn, 2003).

The NLS domain signature at the N terminal of DENV and WNV polymerase domain is also unique to the flavivirus RdRp. The NLS domains form the fingertips (for the bNLS) that provide flexible movement between the domains (Yap et al., 2007) (see chapter 5). The unique NLS sequences are not present either in HCV or BVDV polymerases. Nevertheless, similar fingertips exist in almost all viral RdRps under the genus of pestivirus, hepacivirus, calicivirus or Φ6 bacteriophage polymerase. This special structural feature gives a more spherical and compact appearance to the RdRp, as opposed to a more open and less constrained structure for DNA polymerases. In addition, a motif (motif F) in the fingers domain that was suggested to play an important role in stabilizing the nascent base pair is uniquely present in viral RdRps.

1.8.2 The substrate binding site -Thumb domain

In RdRp, the thumb domain is the most diverse feature among all the known viral RdRps. The thumb domain of Flaviviridae RdRps (such as BVDV) contains 2 times more residues than that of Picornaviridae RdRps and is larger than the thumb domain of other viral RdRps (Fig 1.17). The thumb domain of Flaviviridae RdRps has a β-thumb region (present in BVDV)

or β-hairpin (present in HCV) or priming loop (in DENV and WNV) protrudes from the thumb into the active site that reduces the template channel volume and thus may allows only ssRNA access to the active site during initiation (Fig 1.17). The enzyme RdRp of Φ6 and HCV that use de novo initiation contains C-terminal region that folds into the cavity of the active site (Fig 1.16 & 1.17) (Butcher et al., 2001). The C-terminal folding feature is absent in flavivirus RdRps. In contrast, poliovirus and calicivirus RdRp that use primer dependent initiation lack the β- hairpin or priming loop or C terminal folds to activate and thus have wider template tunnel to accommodate a template-primer complex for initiation [see chapter 5] (Fig 1.17) (Ferrer-Orta et al., 2006b). All known viral RdRp structures so far resolved have a small positively charged tunnel on the back side of the thumb, which may serves as nucleotide diffusion tunnel (Bressanelli et al., 2002; Yap et al., 2007). In addition, structural data on viral RdRp-template complexes (discussed before) show that no interaction occurs between thumb domain and the template/primer. In contrast, the thumb domain of other polymerase families have disclosed important role in positioning, processivity or translocation the duplex DNA (Steitz, 1999).

1.8.3 The catalytic active site

The palm subdomain of RdRps is remarkably similar to the other three classes of polymerases. It appears to be common in catalysis of the phosphoryl transfer reaction (O'Reilly and Kao, 1998). Four polymerases families employ identical two metal ion catalysis (usually Magnesium) at the active site (Steitz, 1998). In this universal mechanism, two divalent metal ions are coordinated by two carboxylate residues, water, the 3'-OH of the growing strand, and the unesterified oxygens of the triphosphate moieties. One metal serves to activate the nucleophilic attack of the 3'OH on the α-phosphate while the other metal assists the leaving of pyrophosphate (Steitz, 1998). From the structural and primary sequence

analysis, four conserved motifs that are (1) involved in phosphoryl transfer and metal ion coordination (motif A and C), (2) nucleotide recognition and binding (motif A and B), and (3) structural integrity of palm domain (motif D) are all well conserved in the viral RdRps palm domain of that from Flaviviridae, Cystoviridae, Togaviradae, Piconaviradae, and Totiviradae. Furthermore, the sequence analysis of the conserved motif A-E sequence analysis shows that these viral RdRps are also highly conserved in HIV RT of RdDps (Bruenn, 2003).

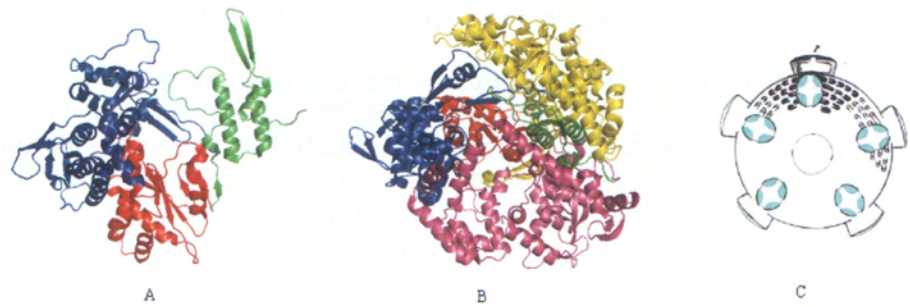


Fig 1. 15.Crystal structure of reovirus polymerase λ3.

(A) Reovirus polymerase λ3 domain, fingers, thumbs and palm domains are colored in blue, green and red respectively. (B) Top view N and C terminal region that surround the polymerase domain, colored in yellow and pink respectively. (C) Reovirus structure shows the position of polymerase λ3 in reovirus core. The dsRNA is packaged in liquid crystalline arrays. The cylindrical protrusions on the outer surface represent the pentamers of the viral capping enzyme. The polymerase complex is represented by the cup shaped object in green color. Adapted from (Tao et al., 2002)

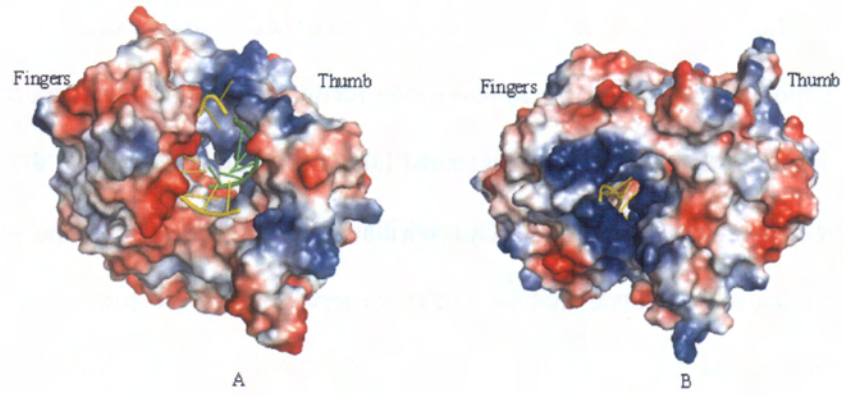


Fig 1.16.Electrostatic surface of FMDV and Φ6 RdRp.

The figure shows the difference between RdRps that use RNA synthesis- primer dependent vs. de novo. (A) FMDV 3D polymerase that use primer dependent initiation has a large and exposed cleft to accommodate the template primer. (B) Φ6 RdRp that use de novo initiation and basically has no room to accommodate primer in addition to the template. It has a C-terminal region that fold into the active site thus stabilizing the initiation complex. The initiation complex is proposed to undergo conformational change in order to proceed to elongation process and to allow the dsRNA product to exit.

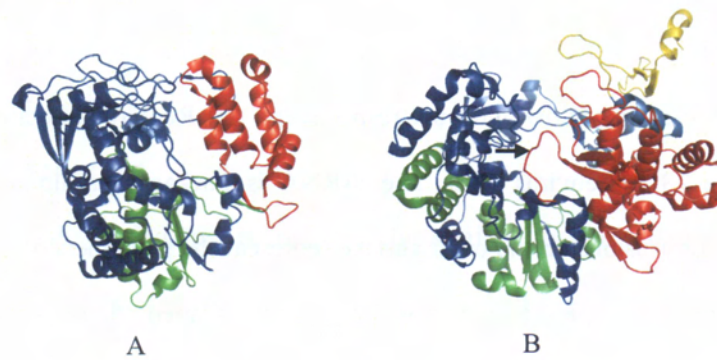


Fig 1.17.Crystal structure of viral RdRp.

(A) Poliovirus 3D polymerase and (B) BVDV RdRp, in cartoon representation. The unique feature of N terminal of BVDV is colored in yellow whereas the special fingertips that exist in almost all the Flaviviridae RdRps, is colored in light blue. The size of thumb domain for

poliovirus 3D polymerase is half of that BVDV RdRp. The arrow indicates the β hairpin of BVDV RdRp protruding into the active site.

1.9 Non Structural Proteins

As discussed earlier and also summarized in Table 1.1 (see later), the DENV encoded polyprotein consists of three structural (C, prM and E) and seven non-structural proteins. An overview of the non-structural proteins is presented below.

1.9.1 NS1

NS1 (46kDa) is a glycoprotein that is essential for virus viability but has no established biological activity (Mason, 1989) It appears either as cell associated or secreted form (Post et al., 1991; Winkler et al., 1988).The NS1 protein can be exported along the secretary pathway to the plasma membrane, where it remains anchored via a glycosyl-phosphatidylinositol group (GPI) (Winkler et al., 1989) or released as a soluble hexamer (sNS1) from infected mammalian cells (Crooks et al., 1994; Flamand et al., 1999). The hexamer sNS1 levels have been linked with viremia levels in secondary DENV2 infection. In addition, secreted NS1 protein was also found circulating in the sera from dengue virus-infected patients throughout the clinical phase of the disease (Alcon et al., 2002; Young et al., 2000). In the *in vitro* dengue infection model, NS1 is translocated into ER and released from the E protein by host signalase (Falgout and Markoff, 1995; Falgout et al., 1989; Chambers et al., 1990b). In the ER, the NS1 forms homodimer and interacts with membrane components to be involved in the early steps of viral replication (Lindenbach and Rice, 1997; Mackenzie et al., 1996). The formation of NS1 dimer is also linked with proteolytic processing at the junction of NS1/2A and this is consistent with the biochemical experiments (Falgout and Markoff, 1995). The NS1 protein comprises hydrophilic amino acids and no transmembrane domain, controversial

with its nature of association with membrane. Dimeric form of NS1 is proposed to create a hydrophobic surface for peripheral association with membranes (Lindenbach and Rice, 2003). Several reports have shown that NS1 is highly immunogenic and induce humoral antibodies (Timofeev et al., 1998; Schlesinger et al., 1993). The antibodies can be protective and kills infected target cells. However, the induced antibodies against NS1 could also plays a role in pathogenesis during dengue virus infection by cross reacting with cell surface antigens or causing host cell death by apoptosis or complement-mediated lysis, through antibody dependent enhancement (ADE) (Falconar, 1997; Lin et al., 2002a; Lin et al., 2003; Lin et al., 2005; Schlesinger et al., 1993). Because of its immunogenic feature, NS1 has been actively pursued as a target in vaccine development (Chaturvedi et al., 2005).

1.9.2 NS2A

NS2A (22 kDa) is a hydrophobic protein that is generated by cleavage at the N-terminal by unknown ER resident peptidase and at C terminal by viral protease in the cytoplasm (Falgout and Markoff, 1995). Thus, NS2A is a membrane spanning protein. NS2Aα is a C-terminal truncated form, that results after cleavage by viral NS2B-3 protease at the Lys 190 (Chambers et al., 1990b; Nestorowicz et al., 1994). Mutated Lys to Ser blocks the production of both NS2Aα and infectious virus thus suggesting that NS2A and NS3 play a role in the viral assembly or release (Kummerer and Rice, 2002). In addition, deletion of NS2A in KUNV replicons is not complemented in trans (Khromykh et al., 2000) and NS3 is required in cis for virus particle formation (Liu et al., 2002), both of which support the notion that NS2A interacts with NS3 during virion replication. Studies in KUN 2A also have shown NS2A interacts with NS3, NS5, and 3'UTR and being localized to vesicle particle (Mackenzie et al., 1998). Thus NS2A has been linked to coordinate the shift between RNA packaging and RNA replication (Mackenzie et al., 1998; Khromykh et al., 2001b).

1.9.3 NS2B

NS2B (14 kDa), a small membrane associated protein, is functions as the cofactor for NS3 protease (Arias et al., 1993; Chambers et al., 1991; Chambers et al., 1993; Jan et al., 1995). The central hydrophilic region of NS2B is required for protease activity and is conserved amongst the flaviviruses whereas the flanking regions are predicted to anchor the host ER membrane (Chambers et al., 1993; Leung et al., 2001). Functional profiling reveals that NS2B-NS3 pro cleaves the NS2B-NS3 linker in cis (Chambers et al., 1993; Wu et al., 2003). In addition, studies on JEV NS2B also have shown its role in rendering the membrane permeability and thereby contribute to virus induced cytopathic effects in infected cells (Chang et al., 1999).

1.9.4 NS3

NS3 (70 kDa) is a multifunctional protein that contains the protease (Chambers et al., 1993; Erbel et al., 2006), helicase (Xu et al., 2005), 5' termini RNA trisphosphatase (Wengler and Wengler, 1993) and NTPase enzymatic activities (Cui et al., 1998; Chen et al., 1997; Cui et al., 1998), that is involved in polyprotein processing and RNA replication (Lindenbach and Rice, 2003). The N-terminal 180 amino acids of NS3 comprises a core serine protease domain, together with the protein NS2B acting as a membrane-anchoring cofactor to form the active protease (Chambers et al., 1993; Wengler and Wengler, 1993; Yusof et al., 2000). The viral NS2B-NS3pro cleaves the viral polyprotein precursor at the junction of NS2A/NS2B, NS2B/NS3, NS3/NS4A, and NS4B/NS5 (Amberg et al., 1994; Yamshchikov and Compans, 1994), as well as at internal sites within C, NS2A, NS3, and NS4A (Chambers and Rice, 1987; Lindenbach and Rice, 2003). In general, the viral protease has specificity for two basic residues (Lys-Arg, Arg-Arg, Arg-Lys) at the canonical P2 and P1 position preceding the

cleavage site, followed by a small amino acid (Gly, Ser or Ala) at the P1' position (Chambers et al., 1990a).

Crystal structure of active DENV 2 and WNV protease domains (NS2B-NS3pro) have been determined (Fig 1.18) (Chambers et al., 1990a; Erbel et al., 2006). The protease belongs to the chymotrypsin family with a classic Ser-His-Asp catalytic triad (Bazan and Fletterick, 1989; Gorbalenya et al., 1989). The crystals of both DENV and WNV protease domains adopt a chymotrypsin like fold with two β barrels and the catalytic triad located in the cleft between the β barrels. Each β barrel is formed by six β-strands (Fig 1.18) (Erbel et al., 2006; Love et al., 1996).The NS2B cofactor domain that surrounds the NS3 protease could stabilize the protease and modulate the protease activity. The NS2B is the cofactor that contributes one β-strand (β-strand 1, NS2B residues 51-57 in DENV) to the N terminal β-barrel. Similarly, NS4A is the cofactor of NS3pro in HCV which is like NS2B of DENV and WNV. Interestingly, the DENV NS3pro structure in the absence of the NS2B cofactor is substantially different from both the WNV and DENV NS2B-NS3pro structures. These differences are observed throughout the entire enzyme and affect the length and location of secondary structure elements. In fact, the NS3pro structure is derived from a refolded protein with weak protease activity. This further supports that the NS2B is crucial for the stabilization of the NS3pro fold (Erbel et al., 2006; Kim et al., 1996). Furthermore, the fold adopted by the C-terminal of NS2B shows remarkable differences between the inhibitor-bound (WNV) and apo-enzyme protease (DENV). The electron density beyond residue 76 is discontinuous in the DENV NS2B-NS3 pro structure, suggesting the C terminal of NS2B adopts multiple conformations in solution. In contrast, WNV crystal structure with inhibitor shows that NS2B forms a belt like structure and interacts with the active site of NS3pro (Erbel et al., 2006). The structure also shows key residues in substrate recognition. Residues

Gly-151, Tyr-161, Tyr-150 and Asp-129 forms the S1 pocket and stabilizes the positively charged side chain of arginine in P1. Whereas residues Asp-82, Gly-83 and Asn-84 of NS2B form the S2 pocket with negative electrostatic environment to attract positive charged of the P2 arginine (Erbel et al., 2006).

The C terminal domain of NS3 consists of ~ 440 amino acids that has been implicated in viral RNA replication, and constitutes a helicase region based on sequence analysis (Lain et al., 1989). The precise biological functions of the NS3 helicase region are unknown, but it has been suggested to assist replication initiation by unwinding RNA secondary structure in the 3' nontranslated region (NTR) (Cui et al., 1998; Chen et al., 1997; Tanner et al., 2003). The flavivirus helicase is a member of the DEAH/D box family within helicase superfamily 2 (Tanner et al., 2003; Gorbalenya and Koonin, 1993). Flavivirus helicase comprises two motifs- Walker A and Walker B, which respectively contains GKS/T and DEAH/D sequences. The DEAH/D in the Walker B motif is believed to interact with divalent cations and NTP substrate. Five other additional conserved motifs are also present in flavivirus helicase, according to the classification of helicases under three superfamilies. Crystal structures for several superfamily 2 of members Flaviviridae helicases have been determined, including DENV2 (Xu et al., 2005), YFV (Mastrangelo et al., 2006) KUN (Wu et al., 2005a) and HCV (Kim et al., 1998) helicase (Fig 1.19). The helicase motifs form an NTPase active site at the interface of two domains (domain I & II) (Xu et al., 2005; Wu et al., 2005a) while a third domain (domain III) also forms a cleft with the other two domains which serve as a binding site for single stranded nucleic acid (Kim et al., 1998). Helicase-nucleic acid co-crystal structures from HCV have shown the RNA template binding position thus suggesting a common mechanism for coupling NTP hydrolysis and strand unwinding for flaviviral helicase (Kim et al., 1998). Despite these details gleaned from the structure, the unwinding

mechanism of flavivirus helicases, their specificity for DNA or RNA duplex, requirement for a 3' and 5' overhang remains unclear. RNA triphosphatase (RTPase) activities of DENV2 helicase was reportedl to be inhibited by Mg^{2+}, high ionic strength and non hydrolysable ATP analogue. A separate study shows that ATPase and RTPase share a common active site (Bartelma and Padmanabhan, 2002). A conserved Q motif upstream of the Walker A was also reported important for ATPase activity of the Powassan flavivirus (Tanner et al., 2003). The RTPase could possibly be involved in dephosphorylation of 5' genomic capping (Wengler and Wengler, 1993). Recently, a complete FL NS3 crystal structure fused to 18 residues of the NS2B cofactor has been determined (Fig 1.20) (Luo et al., 2008). The study reported the relative orientation between the protease and helicase domains is drastically different to that of the single-chain NS3-NS4A molecule from HCV (Fig 1.20). The C-terminal of NS3 HCV participates actively in complex formation with the protease domain which insert into the active site of protease domain to participate in *cis* cleavage that occurs at the NS3-NS4A junction. In contrast, the C-terminal NS3 DENV is far away from the protease active site (Fig 1.20). The authors also proposed that a basic patch located at the surface of protease domain increases the affinity for nucleotides and participate in RNA binding, which unravel higher unwinding activity of the FL NS3 enzyme compared to that of the helicase domain.

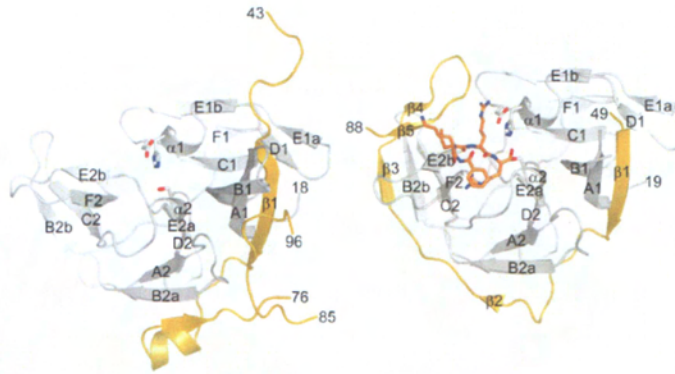


Fig 1.18.Crystal structure of DENV2 and WNV NS2B-NS3pro.

NS2B is represented in yellow, NS3pro in gray. (A) DENV2 NS2B-NS3pro. No electron density was observed for NS2B residues 77-84. (B) WNV NS2B-NS3pro in complex with Bz-Nle-Lys-Arg-Arg-H (in orange). (Erbel et al., 2006).

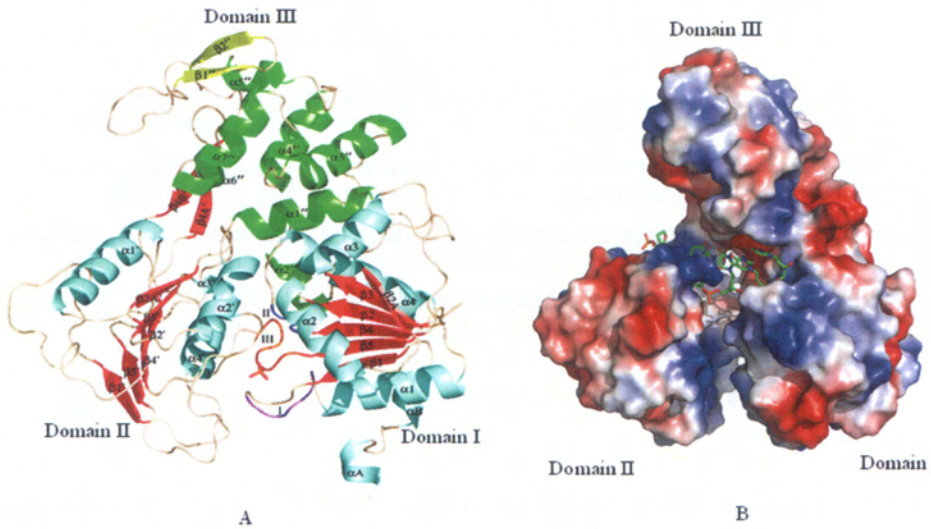


Fig 1.19.Structure of Flaviviridae helicases of DENV2 and HCV.

(A) Overall fold of DENV2 helicase, with each domain labeled respectively. ATP substrate pocket is located between domain I & II. (B) Structure of HCV helicase in complex with dU_8, with the RNA is shown in blue stick, located at the interface created by domain III with two other domains. Figure A and B were adapted respectively from (Xu et al., 2005; Kim et al., 1998).

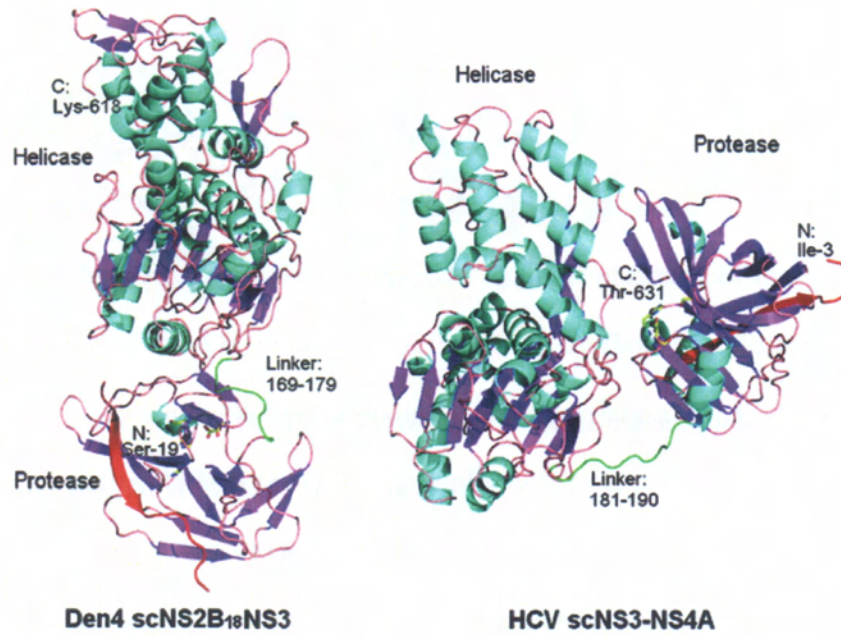


Fig 1.20.Crystal structure of FLNS3 of DENV and HCV in ribbon representation.

Structural comparison of NS2BNS3 DENV4 and NS3NS4A HCV. The helicase orientation is relatively similar for both proteins. The positions of the protease domain relative to the helicase domain in the two structures are clearly different. The cofactors (NS2B and NS4A) are red, and the interdomain linkers are green. Adapted from (Luo et al., 2008).

1.9.5 NS4A

NS4A is a small hydrophobic protein and its function still remains elusive. The N-terminal NS4A is cleaved by the viral NS2B-NS3 protease in the cytoplasm whereas the C-terminal acts as a signal sequence for translocation of NS4B into ER lumen. The C-terminal 23 amino acids signal sequence, also called the 2K fragment, is removed from the N-terminal of NS4B by host signalase in the ER lumen, after viral NS2B-NS3 protease cleavage at the 4A/2K site near the N-terminal of the 2K fragment (Lin et al., 1993). The interaction between YFV NS4A and DENV2 NS1 in trans, localization of KUN NS4A into VP and participation of

NS4A in polyprotein processing have been linked NS4A to be involved in viral RNA replication (Lindenbach and Rice, 1999; Lin et al., 1993). NS4A could function to anchor the replication complex to intracellular membranes. Furthermore, association of NS4A with cellular membranes and induction of membrane reorganization is unknown. The hydrophobic proteins NS2A, NS4A, and NS4B have also been implicated in contributing to the inhibition of the interferon-α/β response (Munoz-Jordan et al., 2005; Munoz-Jordan et al., 2003). Recently, the DENV NS4A membrane topology has been characterized. The study shows the proteolytic removal of 2K fragment can induces cytoplasmic membrane alteration in resembling virus-induced structures and colocalizes with other viral proteins and RNA (Miller et al., 2007).

1.9.6 NS4B

NS4B is a small hydrophobic protein and contains several endoplasmic reticular and cytoplasmic domains that are separated by the transmembrane region. The membrane topology is strikingly similar for flaviviral NS4B despite sequence divergence, suggesting a conserved role of NS4B in viral life cycle (Miller et al., 2006). Deletion studies and trans complementation experiments on NS4B of BVDV and KUNV NS4B were shown to inhibit viral replication (Grassmann et al., 2001; Balint et al., 2005). Role of NS4B in viral replication has been shown in two cases. First, BVDV NS4B was observed to interact with NS3 and NS5A(Qu et al., 2001), Second, HCV NS4B was found to induce morphological changes in the ER membrane (Egger et al., 2002). In addition, KUN NS4B has been shown to localize to the nucleus (Westaway et al., 1997a). Recently, interaction between DENV NS4B and NS3 was identified by Y2H system and pull down assays, suggesting NS4B could enhance the helicase activity by dissociating single strand RNA from NS3 (Umareddy et al., 2006).

Protein	Functions	Structure (PDB code)
C (11 kDa)	As nucleocapsid & binds to 5' & 3' UTR. Internal hydrophobic domain mediates membrane association	X-ray: DENV2 (1R6R) NMR: KUN (1SFK)
prM (26 kDa)	Precursor for of M. Structural protein for maturation of virus.	Cryo-EM: DENV (1P58)
E (53 kDa)	Major virion surface molecule, mediates binding & membrane fusion. In dimer form at neutral pH and as trimer at low pH.	X-ray: DENV with β-OG (1OAM), TBEV (1SVB) Cryo-EM: DENV (1K4R), DENV with glycans (2B6B)
NS1 (46 kDa)	As secreted or non-secreted with membrane associated form. Homodimers involve in RNA replication. Interacts with NS4A in RNA replication. Elicits immune response.	Not available
NS2A (22 kDa)	A transmembrane protein, putative role in virus assembly and release. Also interacts with NS3, NS5 & 3' UTR, possible regulate RNA packaging.	Not available
NS2B (14 kDa)	Cofactor for NS3 protease. Also potentially involves in mediating membrane permeability.	X-ray: DENV2 (2FOM), WNV (2FP7); together with NS3 protease.
NS3 (70 kDa)	Multifunctional protein. N terminal has serine protease role while C termini has NTPase and helicase activities.	X-ray: DENV2 (2BHR), YFV (1YKS), YFV-ADP (1YMF), DENV4 $FLNS2B_{18}NS3$ (2VBC)
NS4A (16 kDa)	Function is not clear. Interacts with NS1& localized at vesicle pocket, possible plays role in replication. Proposed to induce membrane reorganization during infection. Inhibits interferon response.	Not available
NS4B (27 kDa)	Unknown function. Possible to play a role in replication. Localized in nucleus. Proposed enhances helicase activity.	Not available
NS5 (104 kDa)	Multifunctional protein. N terminal has methyltransferase activity (involve in capping), while C termini has RNA dependent RNA polymerase activity. Found localized in the nucleus.	X-ray: DENV3 (2J7U), DENV3-GTP (2J7W), WNV (2HFZ)

Table 1.1.A summary of flaviviral structural and nonstructural proteins

1.10 Antiviral Inhibitor

Early Flaviviridae drug discovery programs were hampered by the lack of animal models and reliable tissue culture replication system to evaluate the antiviral compounds, including DENV. Recently, a dengue fever viremia mouse model that lacks the interferon-α/β receptors

(Schul et al., 2007) and a viral replicon system (Ng et al., 2007) have been established in DENV for drug discovery program. Likewise, HCV replicon systems have been exploited to select for resistance mutations for a variety of inhibitors of key enzymes (Perrone et al., 2007; Blight, 2007). An infectious cell culture system of HCV has also been reported to study the virus entry, virus assembly and release (Zhong et al., 2005). Antiviral chemotherapy targeting viral polymerases have been proven to be efficient for the treatment of human immunodeficiency (HIV), herpes simplex virus and hepatitis B virus (HBV) infection. More than 10 reverse transcriptase inhibitors have been reported and approved for HIV infection treatment (Declercq, 1995). Additionally, a number of HCV RdRp inhibitors are undergoing early phase human clinical trials. See Koch and Narjes, 2007 for review. However, a major difficulty in drug discovery is the side effects and rapid emergence of drug-resistant viral strains.

1.10.1 First antiviral drug discovery- lessons learned in antiretroviral experience

Viral polymerases inhibitors can be divided into nucleoside inhibitors (NI) and non-nucleoside inhibitors (NNI). The NIs serves as substrate for viral polymerases and can be incorporated into the elongative viral RNA. Subsequently, these NIs that act as chain terminator that stop the chain elongation and aborts viral RNA synthesis. Whereas the NNIs mostly act as allosteric inhibitors by targeting a pocket that can prevent protein conformational change. Usually these NNIs are non competitive to the nucleotide substrates.

The first milestone in antiviral drug discovery was that of antiretroviral in HIV. Reverse transcriptase (RT) is a key enzyme which plays an essential and multifunctional role in the replication of the HIV. In the infected cell, the viral RNA is reverse transcribed to produce a double stranded DNA. Reverse transcription is carried out by the RT, using the following

catalytic activities: (1) RNA-dependent DNA polymerization to form an RNA:DNA hybrid (2) RNase H degradation of the RNA strand from RNA:DNA hybrids and (3) DNA-dependent DNA polymerization to form dsDNA. RT has two catalytic domains: a polymerase that copy either RNA or DNA, and an RNase H that degrades RNA strand of RNA-DNA intermediates formed during viral DNA replication. RT consists two subunits, p66 (560 aa) and p51 (440 aa). The p51 subunit appears to provide structural support which allows the p66 subunit to carry out polymerase and RNase H activities. Structural elucidations on RT provide information about the structural changes that occur during polymerization (Majumdar et al., 1988; Reardon, 1992). First, the structure of unliganded RT has a thumb subdomain in the closed conformation (Hsiou et al., 1996). Binding of tempate/primer forms the closed binary complex with the thumb moving to accommodate the nucleic acid (Jacobomolina et al., 1993). Subsequent binding of an incoming dNTP forms the ternary complex that is believed to have the fingers in the open conformation. There is no structural data for this complex. The conversion of the open ternary complex to the closed ternary complex for the formation of phosphodiester bond involves the p66 fingers subdomain movement (Huang et al., 1998). This step is believed a rate limiting step in polymerization.

The first approved antiretroviral drug azidothymidine (AZT) for HIV patients (in the USA and Europe) were nucleoside RT inhibitors (NI) (see below). Highly active antiretroviral therapy (HAART) is the commonly used for therapy, which combines two classes of drugs active against RT (NI) and the NNI with drugs that target the viral protease (Gulick et al., 1997). Although HAART has been shown successfully to suppress viral loads, the virus is not eradicated and persists in latent reservoirs. These reservoirs can serve to replenish the main pool of replicating virus, allowing continued virus evolution and acquisition of resistance (Barre-Sinoussi et al., 2004).

A key biological property of HIV that facilitates the emergence of drug resistance is its capability to evolve rapidly which resulting in genetic diversity. There are two factors that cause the genetic variability: the error prone nature of the HIV polymerase and rapid rate of HIV replication. The generation of HIV genetic variation results in a diverse and related viral population, termed quasi-species, which contains a large number of potentially multidrug resistance (MDR) variants HIV isolates (Barre-Sinoussi et al., 2004). The emergence of MDR HIV isolates has caused an urgent need for new antiretroviral drugs. New drugs should target replication cycle that are not targeted by the current antiviral drugs. The alternative steps in HIV replication: virus entry, integration and maturation, which are the new classes of antiretrovirals target. Some of these antiviral drugs have been approved for clinical use (see Adamson and Freed 2008 for review) (Adamson and Freed, 2008).

In the case of NI resistance, HIV RT circumvent the NI-blocked DNA synthesis so that virus can still replicates. This involves a process to enhance the discrimination between the normal substrate and the analog, which occur either when the NI is incorporated into DNA (incorporation) or after it has been incorporated (excision). In the incorporation process, typical example is steric hindrance caused by mutated residue M184V or M184I that interferes the incorporation of the NI at the step of inhibitor binding or at the step of catalytic incorporation of the inhibitor for 3TC (Neyts and Declercq, 1994). Structural study has suggested that unfavorable interactions between β-branched amino acids (Val or Ile) and the β-L-oxathialone ring of the 3TC (Sarafianos et al., 1999). Interestingly, this steric interference has no effect on the incorporation of normal dNTPs. As shown by many structural studies, fingers subdomains act a clamp to position the template/primer and palm domain is the primer grip that placing the primer terminus at the catalytic site of the polymerase. During nucleotide incorporation, the fingers subdomains is translocated from an

opened to closed complex. Biochemical analysis suggested that the mutation interferes the stable formation of closed complex with 3TCP (Boyer et al., 2000). In the excision mechanism, resistance mutants enhance the excision of the analog after it has been incorporated, notably in 3'-azido-3'-deoxythymidine (AZT) resistance is enhanced excision of AZTMP. AZT resistance mutants unblocks AZTMP-terminated primers by a mechanism that involves nucleophilic attack by a pyrophosphate donor, probably ATP (Boyer et al., 2001). The reaction products are an unblocked primer and a dinucleoside tetraphosphate derived from ATP and the AZTMP at the primer terminus. This reaction allows polymerization to continue and causes resistance to AZT. A set of residues that are responsible for the AZT resistance were identified, these are M41L, K70R, T215F, and K219Q.

Based on various structural, biochemical, and chemical studies, binding of NNIs to HIV-RT were found to block the chemical step of nucleotide incorporation but not prevent the binding of nucleic acid to the enzyme. The NNIs binding pocket is located in the p66 subunit of RT. NNI binding pocket is formed by residues which including aromatic residues Y181, Y188, F227, W229 and Y318 and hydrophobic residues P95, L100, V106, V179, L234 and P236 (Kohlstaedt et al., 1992). Only single residue from the p51 subunit E138 was found to interacts with some NNIs. Although NNIs have diverse chemical structures, they adopt a similar configuration when binding to pocket. The binding of most NNIs cause a large movement of (1) β12- β13- β14 sheet (contains the primer grip) away from the β6- β10- β9 sheet (contains the catalytic residues); and (2) p66 subdomain that in contact with the primer grip, to an upright position. All of the mutations associated with NNIs resistance are located in and around NNIs binding pocket. A set of NNIs resistance residues were identified, these include K103N, Y181C, L100I, V106A, Y188L and G190A (Ren et al., 2002). Structural and biochemical studies have suggested there are three primary mechanisms by which a mutation

can confers resistance to NNIs, these include (1) prevent NNI entry into binding pocket (2)disrupt specific contacts between residues that form the pocket and inhibitor and (3) resistance mutants have effect on the conformation or size of the pocket (Ren et al., 2002; Auwerx et al., 2004). It is common that mutation that affects NNI binding could be due to more than one mechanism.

In the case of Flaviviradae, antiviral drug design efforts such as NI of HCV tend to be broad acting. They are active in inhibiting flaviviruses including WNV, DENV, and YFV (Eldrup et al., 2004). The viral inhibitors discussion below is focused on the Flaviviridae polymerase.

1.10.2 Nucleoside inhibitors

Nucleoside inhibitors are prodrugs, which have to be first converted to the active triphosphate via cellular kinases. These phosphorylated NI's target the active site of polymerase and cause inhibition by chain termination or introduction of drastic mutations that scramble the genome. The efficacy of NI depends on the NTP's affinity to the polymerase, the ratio of NTP analog to the endogenous NTP and incorporation rate to chain elongation. Poor cell permeability, poor conversion to the triphosphate and unwanted metabolic side-products also affect the NIs activity. Toxicity is also another issue that limits the potency of NIs. In HCV RdRp NI, small changes in the 2' or 4' position of the sugar were sufficient to convert a natural substrate into a chain terminator (Olsen et al., 2004).

Example of nucleoside inhibitors

A series of 2'- modified ribonucleosides were found to inhibit HCV *in vivo* and *in vitro*, independently by Merck/Isis and Idenix/Novartis (Fig 1.21). The inclusion of a 2'-C-methyl group either to adenosine or guanosine yields potent HCV compounds (Eldrup et al., 2004;

Wu et al., 2005b). Modifications on the ribose with ethyl at 2'-methyl position, changes of the stereochemistry of the methyl group or regional chemistry at 2' or 3' position lead to a complete loss of activity. These facts indicate that the chemical space variable for modifying nucleoside for investigation is narrow. This highlights the difficulty in performing lead optimization on the nucleotide scaffold and and also brings up issues concerning the overcrowded chemical space for patents. Both the 2'-C-methyl purine nucleosides (adenosine and guanosine) show IC50 of 1.9 and 0.13 μM, respectively (Wu et al., 2005b). The 2'-C-methyl purine nucleosides presumably impair the elongation by 3'-OH disorientation that is required for new incoming nucleotide incorporation or mis-alignment of α-phosphate of the incoming nucleotide. The 2'-C-methyl adenosine is more potent than guanosine in replicon cells, suggesting the cellular uptake or intracellular metabolism for the latter is less efficient. Nevertheless, the adenosine nucleoside exhibits high plasma clearance and show a lack of oral bioavailability. A 2'-C-methyl purine nucleosides resistant clone was also identified from replicon system, which is at residue Ser-282 (Migliaccio et al., 2003). The mutant S282T reduces the 2'-C-methyl purine nucleosides binding affinity by occupying the space that occupied by the 2'-C-methyl. Further modification efforts led to discovery of 7-Deaza-2'-C-methyl adenosine, which has good pharmacokinetic profile, oral bioavailability and moderate plasma clearance (Koch and Narjes, 2007). This compound also shows antiviral activity against related flaviviruses such as BVDV, WNV and YFV, but did not inhibit human DNA and RNA polymerases, making it an attractive target for clinical development. In addition, nucleoside inhibitor NM283 is a valine prodrug of 2'-C-methyl-cytidine (NM107) (Fig 1.21) (Pierra et al., 2006). Proof of concept of NM283 was achieved in Chimpanzees, with one $\log_{10}$ viral reduction. PK analysis also shows that NM283 is efficiently converted to 5'-triphosphate *in vivo* with a long half life of 13.8 hrs. A combination of NM283-IFN treatment on HCV infected patients are undergoing at phase IIb to evaluate treatment-naive patients and

also patients who failed to respond to PEG-IFN and Ribavarin treatment (Koch and Narjes, 2007). Drug resistant S282T clone was also identified and reduces 30-fold potency for NM107. However, NM283 was terminated due to lack of efficacy.

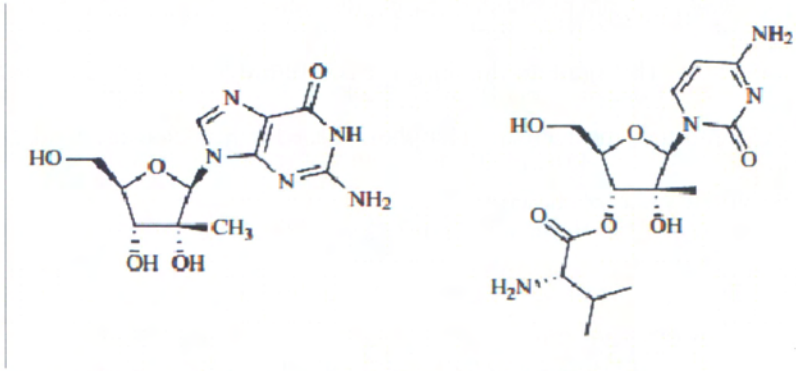


Fig 1.21.Structure of 2'-C-methyl-guanosine(left) and NM 283 (right).

1.10.2 Allosteric inhibitor

So far, crystal structure for Flaviviridae polymerase-inhibitors complexes is only available for HCV. Several HCV RdRp-NNI crystal structures revealed two sites at the thumb domain and a site in the palm-thumb interface that could bind three distinct allosteric sites. For the thumb allosteric site, one of the conserved site is the region where thumb interacts with fingertips (NNI site I) and another site lies at the base of the thumb domain (NNI site III) (Fig 1.22).

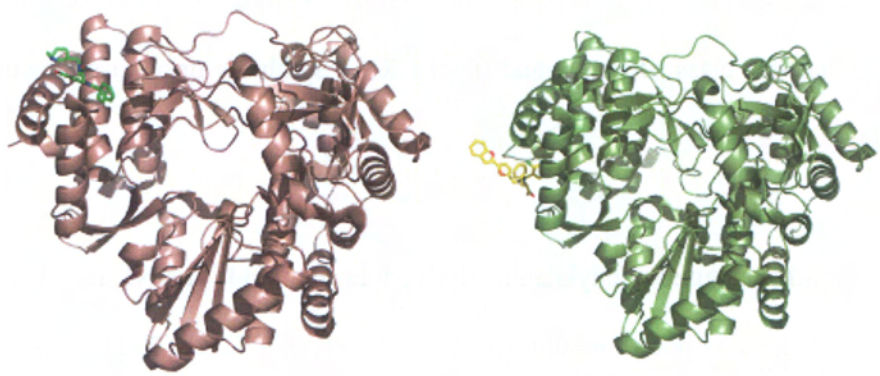


Fig 1.22.Crytal structures for HCV RdRp-inhibitor.

(Left) HCV-RdRp-indole based inhibitor crystal structure. The inhibitor binds on the surface of thumb domain which interacts with the fingertips. This site is referred as NNI site I. The binding of NNI at site I is suggested to interfere the concerted movement of fingertips during RNA replication. (Right) A phenylalanine based inhibitor bound at thumb domain, remotely from active site (~35 Å). This inhibitor binding site is referred as NNI site III. See text for the proposed site III NNI inhibition mechanism. Thiophene-based non nucleoside inhibitors are also found to bind at the similar site (not shown).

Example of allosteric inhibitor

Benzimidazoles based compounds have been identified to be potent in inhibiting HCV RdRp (Dhanak et al., 2002). The binding of inhibitor at NNI site I was revealed by crystal structure (Tedesco et al., 2006). Initial SAR on benzimidazole series revealed that the minimum core required for activity is a benzimidazole with a cylohexyl at N1, aryl group at the C2 and carboxylate at C5. Japan Tobacco focused on the right hand side of the compound (Fig 1.23a) and changed the benzyl group to biphenyl group yielding a potent compound (JTK-109) (Fig 1.23b). The JTK-109 and an analog (JTK-003) were in early phase clinical trials were discontinued for development. A single mutation P495L in HCV RdRp was identified in replication drug resistance studies and this pocket is suggested to be too small to accommodate the compound from structural data (Migliaccio et al., 2003). Kinetic analysis show that this class of compound blocks RNA synthesis prior to elongation (Tomei et al., 2004).

The N,N-disubsituted phenylalanine derivatives that binds at NNI site III were discovered by Shire Biochem as potent inhibitor in HCV RdRp (Fig 1.24) (Reddy et al., 2003). However, this inhibitor class is inactive in the replicon system, even though the inhibitor binding is

supported by the crystal structure (Fig 1.22, right) (Wang et al., 2003). As an extension of the phenylalanine series, inhibitors based on 5-phenyl-thiophene scaffold such as the sulfonamide were reported (Chan et al., 2004). Potency was improved by conversion of the amide to the carboxylic acid and SAR around the sulfonamide group. This class of inhibitor is active in the replicon assay. The structures of thiophene sulfonamides bound to HCV RdRp revealed that these inhibitors have common binding characteristic and are similar to the phenylalanine-based inhibitors at NNI site III. All these inhibitors induce perturbations of the pocket by side chain displacement of Met-423 and Leu-497 (Tomei et al., 2003). Movement of residue Leu-497 also causes conformational change of Pro-496 and Arg-505 which these residues lie on the helix T. The shift in helix T compromises the GTP binding site integrity (Tomei et al., 2003). In addition, the phenylalanine derivatives inhibitors do not inhibit human DNA polymerase α, β or γ (IC50 > 100 μM).

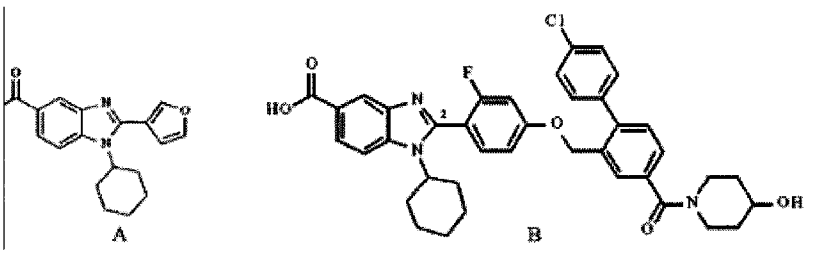


Fig 1.23.Benzimidazoles inhibitor.

(A) The basis structure of benzimidazoles. (B) JTK-109.

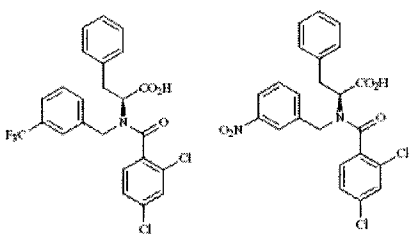


Fig 1.24.Phenylalanine based inhibitors

1.11 Aims of the thesis

Resurgence and emergence of dengue in recent years show that there is a need to prepare for the epidemic surge with specific antiviral drugs. At the present moment, there is no effective therapy for dengue. NS5 is a multifunctional protein that plays a central role in viral replication, and is endowed with MTase and polymerase activities that are located in separate domains. Selective inhibitors against human immunodeficiency (HIV), herpes simplex virus and hepatitis B virus (HBV) have been approved as drugs for these viral related treatments. Therefore, NS5 represents a very attractive target for the development of specific antiviral inhibitors. Several Flaviviridae polymerase structure such as those of BVDV and HCV have been presented. However, at the inception of this study there was no polymerase crystal structure of flaviviruses, including DENV. While NS5 shares several common characteristics with HCV and BVDV in viral life cycle, such as the polymerase employs common polymerization strategy, there are many differences in the function of viral replication. As discussed earlier, dengue NS5 contains MTase domain and is expressed as single polypeptide with polymerase whilst the MTase does not exist in HCV and BVDV. In addition, dengue virus initiates translation with a capping process that is mediated by MTase, instead of IRES (internal ribosomal entry site) as employed in HCV. Other marked difference between HCV and DENV is the presence of nuclear localization signal in DENV NS5. Therefore, there is a motivation and a demand to understand the DENV RdRp 3D structure to shed light on its role and function as well as provides insight for rational antiviral drug design.

Specifically, the goals of this thesis are:

(1) To explore soluble NS5 proteins expression strategy for functional and structural studies including the use of specific antibodies generated as part of this study in order to stabilize the protein for crystallization (chapter 3).

(2) To investigate a multi-step strategy in order to obtain the correct RdRp crystal condition for high resolution studies and to employ X-ray crystallographic screening to discover ligands (chapter 4)

(3) To obtain insight into crystal structure of DENV RdRp and understand the role of divalent metal ion from structural basis (chapter 5).

CHAPTER 2

MATERIAL AND METHODS

Table 2.1

No	Constructs (aa)	Plasmid	Primers (5'-3')	Restriction sites	Comments /source
1.	341-900 DENV2	pET 32a+	F- cacggatccgatgagaacaagtggaag R- ctgcaggtcgacttaccacaagactcctgcctc	Bam HI Sal I	DENV2 TSV01
2.	341-900 DENV2	pET 21b+	F-cacggatccgatggcaatgacagacacgactcc R-gtgctcgagccacaagactcctgcctcttc	Bam HI Xho I	DENV2 TSV01
3.	341-900 DENV2	pFastBac HTB	F- cacggatccgatgagaacaagtggaag R- ctgcaggtcgacttaccacaagactcctgcctc	Bam HI Sal I	DENV2 TSV01
4.	415-900 DENV2	pET 32a+	F-cacggatccgatgagaacaagtggaag R-ctgcaggtcgacttaccacaagactcctgcctc	Bam HI Sal I	DENV2 TSV01
5.	415-900 DENV2	pET 21b	F-cacggatccggatgagaacaagtggaagtc R-gtgctcgagccacaagactcctgcctcttc	Bam HI Xho I	DENV2 TSV01
6.	415-900 DENV2	pFastBac HTB	F- cacggatccgatgagaacaagtggaag R- ctgcaggtcgacttaccacaagactcctgcctc	Bam HI Sal I	DENV2 TSV01
7.	1-900 DENV2	pFastBac HTB	F-cacggatccggaactggcaacacaggagag R-ctgcaggtcgacttaccacaagactcctgcctc	Bam HI Sal I	DENV2 TSV01
8.	1-900 DENV2	pBakPak 9	(gift byDr. Subhash Vasudevan)		DENV2 TSV01
9.	1-900 DENV2	pET 32a+	F-cacggatccggaactggcaacacaggagag R-ctgcaggtcgacttaccacaagactcctgcctc	Bam HI Sal I	DENV2 TSV01
10.	48-359 DENV2	pGEX 4T1	F-cacgatccgcaggagaaacggaccatcac R-ctgcaggtcgacttatttctctttgaaaac	Bam HI Sal I	DENV2 TSV01
11.	264-359 DENV2	pGEX 4T1	F-cacggatccgcaaacatcggaattgaaagt R-ctgcaggtcgacttatttctctttgaaaac	Bam HI Sal I	DENV2 TSV01
12.	48-866 DENV2	pET 32a+	F-cacggatccgcaggagaaacggaccatcac R-ctgcaggtcgacttacctgcttgttagcccaat	Bam HI Sal I	DENV2 TSV01
13.	48-900 DENV2	pET 32a+	F-cacggatccgcaggagaaacggaccatcac R-ctgcaggtcgacttaccacaagactcctgcctc	Bam HI Sal I	DENV2 TSV01
14.	297-866 DENV2	pGEX 4T1	F-cacggatccgcagaccacccatacaaaacg R-ctgcaggtcgacttacctgcttgttagcccaat	Bam HI Sal I	DENV2 TSV01
15.	264-866 DENV2	pET 32a+	F-cacggatccgcaaacatcggaattgaaagt R-ctgcaggtcgacttacctgcttgttagcccaat	Bam HI Sal I	DENV2 TSV01
16.	359-866	pET 32a+	F-cacggatccgcagtggacacgagaacccaa	Bam HI	DENV2

	DENV2		R-ctgcaggtcgacttacctgcttgttagcccaat	SaI I	TSV01
17.	1-296 DENV2	pQE 30	F-cacggatccggaactggcaacacaggagag R-ctgcaggtcgacttattggtcatagtgccatga	Bam HI SaI I	DEN2 TSV01
18.	1-296 DENV2	pQE 30	(Gift by Dr Bruno Canard)		DENV2 NGC
19.	273-900 DENV1	pET 15b+	F-cgaaggctcgagaacctagatatcattggccag R-ggacgtagatctctaccagagtgccccttcgagatc (Gift by Dr Chen Yen Liang)	Xho I BgI II	DENV1
20.	273-900 DENV2	pET 15b+	F-cgaaggctcgagaatttagacataattgggaaaaga R-cgcggatccctaccacaagactcctgcctcttcc (Gift by Dr Chen Yen Liang)	Xho I Bam HI	DENV2 TSV01
21.	273-900 DENV3	pET 15b+	F-ccacgcgtcgacaacatggatgtcattggggaaag R-ccggaattcctaccaaatggctccctccgactcctc (Gift Dr Chen Yen Liang)	Sal I EcoRI	DENV3
22.	273-900 DENV4	pET 15b+	F-cgaaggctcgaggacatgacaatcattgggagaag R-ccggaattcctacagaactccttcactctcgaaag (Gift by Dr Chen Yen Liang)	Xho I EcoRI	DENV4
23.	273-900 DENV1	pDEST 14	F- cgatccgctagcaacctagatatcattggccaga R-caatagcggccgcctaccagagtgccccttcgagatca (Gift by Dr Chen Yen Liang)		DENV1
24.	273-900 DENV2 NGC	pET 28a+	(Gift by Dr Chen Yen Liang)		DENV2 NGC
25.	1-906 WNV	pET 28a+	F-gcgcggcagccatatgggtggagc R-ctcgagtgcggccgcctacaaaacagt	Nde I Not I	WNV

Table 2.1.Detail of NS5 constructs.

Constructs design, primers used for PCR amplification and cloning, expression vectors and restriction digestion sites for respective protein constructs used in this work. Two viral strains of DENV2 are used in this study: the Townville strain (TSVO1, Genbank accession AY037116) or the New Guinea C strain (NGC, Genbank accession M29095). NS5 related genes from other viral strains are as following: DENV1 (Genbank accession EU848545), DENV3 (Genbank accession AY662691), DENV4 (Genbank accession AY947539) and WNV (Genbank accession: AY532665).

Table 2.2

No	Construct (aa)	Plasmid	Tag position	Expression host	Expression conditions	Antibiotics	Cleavage protease
1	341-900 DENV2	pET 32a+	N-terminal Trx &His_5	BL21(DE3)RIL	16°C, 16-18h, IPTG induction	Amp & Chl	Enterokinase
2	341-900 DENV2	pET 32a+	N-terminal Trx &His_5	Origami (DE3)	13°C, 16-18h, IPTG induction	Amp & Chl	Enterokinase
3	341-900 DENV2	pET 21b+	N-terminal His_5	BL21(DE3)RIL	16°C, 16-18h, IPTG induction	Amp & Chl	Enterokinase
4	341-900 DENV2	pFastBac HTB	N-terminal His_5	H5 cells	28°C, 72h, Baculovirus infection	Gentamicin	Enterokinase
5	415-900 DENV2	pET 32a+	N-terminal Trx &His_5	BL21(DE3)RIL	16°C, 16-18h, IPTG induction	Amp & Chl	Enterokinase
6	415-900 DENV2	pET 32a+	N-terminal Trx &His_5	Origami (DE3)	16°C, 16-18h, IPTG induction	Amp & Chl	Enterokinase
7	415-900 DENV2	pET 21b+	C-terminal His_5	BL21(DE3)RIL	16°C, 16-18h, IPTG induction	Amp & Chl	N.A
8	415-900 DENV2	pFastBac HTB	N-terminal His_5	H5 cells	28°C, 72h, Baculovirus infection	Gentamicin	Enterokinase
9	1-900 DENV2	pfastBac HTB	N-terminal His_5	H5 cells	28°C, 72h, Baculovirus infection	Gentamicin	Enterokinase
10	1-900 DENV2	pBakPak 9	N-terminal Trx &His_5	H5 cells	28°C, 72h, Baculovirus infection	Gentamicin	Thrombin
11	1-900 DENV2	pET 32a+	N-terminal Trx &His_5	BL21(DE3)RIL	16°C, 16-18h, IPTG induction	Amp & Chl	Enterokinase
12	48-359 DENV2	pGEX 4T1	N-terminal GST	BL21(DE3)RIL	16°C, 16-18h, IPTG induction	Amp & Chl	Thrombin
13	264-359 DENV2	pGEX 4T1	N-terminal GST	BL21(DE3)RIL	16°C, 16-18h, IPTG induction	Amp & Chl	Thrombin
14	48-866 DENV2	pET 32a+	N-terminal Trx &His_5	BL21(DE3)RIL	16°C, 16-18h, IPTG induction	Amp & Chl	Enterokinase
15	48-900 DENV2	pET 32a+	N-terminal Trx &His_5	BL21(DE3)RIL	16°C, 16-18h, IPTG induction	Amp & Chl	Enterokinase
16	297-866 DENV2	pGEX 4T1	N-terminal GST	BL21(DE3)RIL	16°C, 16-18h, IPTG induction	Amp & Chl	Thrombin
17	264-866 DENV2	pET 32a+	N-terminal Trx &His_5	BL21(DE3)RIL	16°C, 16-18h, IPTG induction	Amp & Chl	Enterokinase
18	359-866 DENV2	pET 32a+	N-terminal Trx &His_5	BL21(DE3)RIL	16°C, 16-18h, IPTG induction	Amp & Chl	Enterokinase

Continue Table 2

No	Construct (aa)	Plasmid	Tag position	Expression host	Expression conditions	Antibiotics	Cleavage protease
19	1-296 DENV2	pQE30	N-terminal His_5	M15 or SG	30°C, 16-18h, IPTG induction	Amp & Kan	N.A
20	1-296 DENV2NGC	pQE30	N-terminal His_5	M15 or SG	16°C, 16-18h, IPTG induction	Amp & Kan	N.A
21	273-900 DENV1	pET 15b+	N-terminal His_5	BL21(DE3)RIL	16°C, 16-18h, IPTG induction	Amp & Chl	Thrombin
22	273-900 DENV2	pET 15b+	N-terminal His_5	BL21(DE3)RIL	16°C, 16-18h, IPTG induction	Amp & Chl	Thrombin
23	273-900 DENV3	pET 15b+	N-terminal His_5	BL21(DE3)RIL	16°C, 16-18h, IPTG induction	Amp & Chl	Thrombin
24	273-900 DENV4	pET 15b+	N-terminal His_5	BL21(DE3)RIL	16°C, 16-18h, IPTG induction	Amp & Chl	Thrombin
25	273-900 DENV1	pDEST 14	C-terminal His_5	BL21(DE3)RIL	16°C, 16-18h, IPTG induction	Amp & Chl	Thrombin
26	273-900 DENV2 NGC	pET 28a+	N-terminal His_5	BL21(DE3)RIL	16°C, 16-18h, IPTG induction	Amp & Chl	Thrombin
27	1-906 WNV	pET 28+	N-terminal His_5	BL21(DE3)RIL	16°C, 16-18h, IPTG induction	Amp & Kan	Thrombin

Table 2.2.Expression conditions

Expression conditions and host cells for respective NS5 related proteins that were used in this study. Amp: Ampicilin, Chl: Chloramphenecol.

Table 2.3

Protein	**Purification buffer**	**Column**
DENV2 FL NS5 (1-900)	Metal affinity (buffer A). 20 mM Tris HCl pH 7.5, 0.5 M NaCl, 0.01% CHAPS, 5 mM β-ME, 10% glycerol and 50 mM imidazole	HisTrap HP column
	Ion Exchange (buffer B). 50 mM Hepes pH 7.0, 80 mM NaCl, 5 mM BME, 10% glycerol, 1 mM EDTA	Source 15S or HiTrap SP FF column
	Gel filtration (buffer C). 20 mM Tris HCl pH 6.8, 0.5 M NaCl, 10 mM BME, 1 mM EDTA and 0.1% CHAPS at 5-10 mg/ml or 20 mM Tris HCl pH 7.5, 10% glycerol, 0.2 mM EDTA, 10 mM β-ME and NaCl (range from 0.2 M-0.5 M) at 5-10 mg/ml	HiPrep 16/26 Superdex 200
WNV FL NS5 (1-900)	Metal affinity (buffer A). 20 mM Tris HCl pH 7.5, 0.5 M NaCl, 0.01% CHAPS, 5 mM β-ME, 10% glycerol and 50 mM imidazole	HisTrap HP column
	Ion Exchange (buffer B). 50 mM Hepes pH 7.0, 80 mM NaCl, 5 mM BME, 10% glycerol, 1 mM EDTA	Source 15S
	Gel filtration (buffer C). 20 mM Tris HCl pH 6.8, 0.5 M NaCl, 2 mM BME, 1 mM EDTA and 0.1% CHAPS at 10-13 mg/ml	HiPrep 16/26 Superdex 200
DENV 3 273-900 & DENV 1 273-900	Metal affinity (buffer A). 20 mM Tris HCl pH 7.5, 0.5 M NaCl, 0.01% CHAPS, 5 mM β-ME, and 10% glycerol	HisTrap HP column
	Ion Exchange (buffer B). 50 mM Mes pH 6.8, 50 mM NaCl, 10 mM BME, 10% glycerol and 1 mM EDTA	Source 15S
	Gel filtration (buffer C). 20 mM Tris HCl pH 6.8, 0.25 M NaCl, 2 mM BME, 1 mM EDTA and 0.1% CHAPS at 10 mg/ml	HiPrep 16/26 Superdex 200
48-359	GST (buffer A). PBS pH 7.3 (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4)	GSTPrep FF 16/10
	Ion Exchange (buffer B). 50 mM Tris HCl pH 7.0, 50 mM NaCl, 5 mM BME, 10% glycerol, 1 mM EDTA	HiTrap SP FF column
	Gel filtration (buffer C). 50 mM Bicine pH 7.5, 0.25 M NaCl, 2 mM BME, 1 mM EDTA and 10% glycerol. (not tested for final concentration)	Superdex 200 10/300GL

Continue Table 3

Protein	**Purification buffer**	**Column**
1-296	Metal affinity (buffer A). 20 mM Tris HCl pH 7.5, 0.5M NaCl, 0.01% CHAPS, 5 mM β-ME, and 10% glycerol	HisTrap HP column
	Ion Exchange (buffer B). 50mM Bicine pH 7.5, 5mM BME, 10% glycerol, 1mM EDTA	HiTrap SP FF column
	Gel filtration (buffer C). 50mM Bicine pH 7.5, 0.8 M NaCl, 2 mM BME, 1mM EDTA and 10% glycerol at 25mg/ml	Superdex 200 10/300GL

Table 2.3.Purification buffer.

Details for large scale protein purification buffers that are described in material and methods. This table also contains respective column used in the purification process.

Table 2. 4

Crystallization kit	**Reference**	**Condition**
Index	Hampton research	96
Salt RX	Hampton research	96
Crystal Lite	Hampton research	50
Crystal Screen	Hampton research	96
Peg/ Ion	Hampton research	48
Wizard I & II	Emerald Biosystems	96
Wizard III	Emerald Biosystems	50
JBScreen HTS I	Jena Bioscience	96
JBScreen HTS II	Jena Bioscience	96
Peg Suite	Nextal, Qiagen	96
Peg II Suite	Nextal, Qiagen	96

Continue Table 2. 4

Crystallization kit	Reference	Condition
Ammonium Sulfate grid screen	Self Make	24
2-Methyl-2,4-pentanediol Grid screen	Self Make	24
PEG 400 Grid screen	Self Make	24

Table 2.4.Crystallization assays conditions.

Screening kits used for DENV2 FLNS5 (both TSVO1 and NGC strain), DENV 1 & DENV 3 RdRp and WNV FL NS5 proteins.

2.0 Molecular cloning

All the NS5 related were amplified by PCR and subsequently cloned into their plasmids at respective restriction sites (Table 2.1). Individual primers, restriction sites, vectors and original templates used in cloning are listed in Table 2.1. In general, Pfu polymerase (Stratagene) was used in PCR amplification with the following design: denaturation [94°C 2 min, 94°C 30 sec], annealing [follow respective T_m of primer sets, or in brief 55°C], extension [72°C 3 min]. This cycle go for 30 rounds before an extension for 72°C 3 min and is followed by 4°C for cooling down and harvest the amplified fragments. The use of 1-5 % v/v DMSO was included in the reaction for difficult amplification. The PCR products were either purified with the QIAquick PCR purification kit (Qiagen Inc) or gel purified using manufacturer's instruction (Qiagen Inc). The amplified fragments were double digested by respective restriction enzymes (Table 2.1) and gel purified, before it was ligated into targeted vectors using T4 ligase (Roche) at 16°C for overnight. The ligated products were transformed into XL1-Blue (cloning hosts) for colony selections and positive clones were confirmed by sequencing.

2.1 Recombinant DNA transfection

The construction bacmids and transfection works were done for construct 3, 6, and 7, table 2.1 using kits from Invitrogen. Briefly, recombinant pFastBac plasmids were transformed into DH10Bac. Once the pFastBac plasmid is transformed into DH10Bac cells, transposition occurs between the mini-Tn7 element on the pFastBac vector and the mini-attTn7 target site on the bacmid to generate a recombinant bacmid. This transposition reaction occurs in the presence of helper plasmid in DH10Bac cells. White clones containing the recombinant bacmids were selected on LB plates with 50 μg/ml kanamycin, 7 μg/ml gentamicin, 10 μg/ml tetracycline, 100 μg/ml blue gal and 40 μg/ml IPTG. The recombinant Bacmid DNA was isolated and transfected into Sf9 insect cells with Cellfectin as described in the manufacturer's manual (Invitrogen). Recombinant baculovirus were harvested from supernatant and titrated by viral plaque assay. In addition, the pBakpak9 NS5 recombinant virus was provided by Dr. Subhash Vasudevan (construct 1, Table 2.1).

2.2 Expression of NS5 from Sf9 insect cells

The recombinant baculovirus was amplified first from the monolayer cultures of Sf9 cells (Grace's medium with FBS) and second from the Sf9 cells (SFM 900 medium without FBS), at MOI of 0.1. The harvested virus was then used to infect High five cells for expression, at MOI of 1-10 for 36 -72 h (refer to manufacturer's manual, Invitrogen). For the case of pBakpak9 NS5 recombinant virus, the high five cells were infected at MOI of 1 and incubated for 72h before harvested for lysis and purification.

2.3 Protein expression from *E.coli*

Recombinant plasmid carrying the gene encoding NS5 constructs were transformed into respective *E. coli* strains and expressed as listed in Table 2.2. The transformed cells were

grown at 37°C in LB medium containing respective antibiotics (Table 2.2) until OD_{600} of 0.6-0.8. Protein expression was induced at respective temperature (Table 2.2) by adding isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.4-0.5 mM (Table 2.2). Cells were harvested by centrifugation at 8,000 × g for 10 min at 4°C and the cell pellet was stored at -80°C. Protein solubility was estimated visually from SDS Polyacrylamide gel analysis.

Fab antibody expression was carried out essentially as discussed in manufacturer' manual (Morphorsis). Briefly, plasmid pMORPH®9_MH that contains the Fab genes (Morphorsis) were transformed into *E.coli* TG1 F-. The cells were spreaded on Chloramphenecol/glucose agar plate and incubated overnight at 37ºC. A 10 ml overnight culture growed in 2x TY/ Chloramphenecol/ 1% glucose medium at 30ºC was prepared. A volume of 2.5 ml from the overnight culture was inoculated into 750 ml 2x TY medium that contains Chloramphenecol/0.1, 1% glucose and incubated at 30ºC in a shaker until an OD_{600nm} of 0.5 was reached, before the culture was induced with 0.75 mM IPTG for overnight expression. The culture was harvested on the next day.

2.4 Protein purification

2.4.1 General lysis

Cells resuspended in a lysis buffer A (Table 2.3) supplemented with an EDTA-free protease inhibitor tablet (Roche) were lysed by sonication, and then lysate was clarified by centrifugation at 30, 000 x g for 30 min at 4°C. Other proteins prepared in this work duting the course of my PhD studies were lysed in similar buffer A of DENV2 FLNS5.

2.4.2 Small scale metal affinity purification

The clarified lysate (3-10 ml) were transferred directly to a 15ml falcon tube containing 500 µl nickel beads agarose (Qiagen) left to bind the beads with rotation. After 1-2 h at 4°C, the tube was centrifuged at 500 x g for 10 min. The supernatant was discarded and the beads were washed three times with wash buffer to remove unbound proteins followed by centrifugation at 500 x g for 5 min. His tagged proteins were eluted with the wash buffer containing 0.5 M imidazole. Small scale GST tagged protein purification was performed similarly.

2.4.3 Large scale Metal affinity purification.

For large scale purification (see Table 2.2 for His tagged proteins), the lysate supernatants were first purified by metal affinity using a HisTrap HP 5ml column (GE Healthcare) equilibrated with buffer A DENV2 FLNS5 (Table 2.3). However, the subsequent steps of purification and buffers were optimized for each individual protein. A full description of DENV 3 RdRp purification procedure for crystallization studies is described in Yap et al., 2007.

The following general purification method applied for proteins that were prepared for structural studies in this work-DENV2 FLNS5, WNV FLNS5, DENV 1 and DENV 3 RdRp, and MTase DENV2 (refer to protein 10, 25, 24 and 22 respectively in Table 2.2). DEN RdRp representing the catalytic domain of the polymerase starts at residue 273-900. After lysis, all supernatants were flowed through the HisTrap HP 5 ml column. Unbound proteins (except for DEN2 FLNS5 purification) were washed away sequentially with five column volumes each of buffer A supplemented with 25 mM and 125 mM imidazole respectively. Proteins were then eluted using a linear gradient of imidazole from 125 to 500 mM. For DENV2

FLNS5, unbound proteins were washed away with five column volumes of buffer A. Proteins were then eluted using a linear gradient of imidazole from 50 to 250 mM.

2.4.4 Ion exchange purification

Fractions containing the protein were pooled and dialyzed overnight against buffer B (table 2.3). After dialysis, proteins retaining the N-terminal His tag were subjected to cation-exchange chromatography. Proteins were eluted using a linear gradient from 0.15 to 1.5 M NaCl. Fractions that contain the proteins were concentrated and subjected to gel filtration purification.

2.4.5 Thrombin digestion

Alternatively, after elution from the HisTrap HP column, the N-terminal tagged proteins (DENV 3 RdRp, DENV 2 FLNS5 and WNV NS5) were treated with thrombin (1 U thrombin per 100 mg of protein; GE Healthcare) in 20 mM Tris HCl pH 7.0-8.0, 50 mM NaCl and 5 mM β-ME overnight at 4°C to remove the His tag. The thrombin-treated proteins were dialyzed in buffer B, followed by purification using the Source 15S ion exchange chromatography.

2.4.6 Gel Filtration

As a final polishing step, proteins were further purified with HiPrep 16/26 Superdex 200 in buffer C. The eluted monomeric proteins were concentrated to respective concentration (Table 2.3) in mg/ml and used in crystallization assays (Table 2.2 & 2.4).

2.4.7 GST tagged protein large scale purification.

For GST tagged NS5 48-359, proteins was lysed in buffer A supplemented with 1mM PMSF and was first purified by GSTPrep FF 16/10 Column (GE Healthcare). Proteins were eluted using 10 mM reduced glutathione. Fractions containing the protein were pooled and diluted with buffer B. After dilution, proteins were subjected to cation-exchange chromatography using a HiTrap SP FF column (GE Healthcare). Proteins were eluted using a linear gradient from 0.15 to 1.5 M NaCl in buffer C and further purified by gel-filtration chromatography Superdex 200 10/300GL.

2.4.8 Fab antibody purification

The buffer used for Fab antibody purification was similar used for DENV 2 FLNS5. The proteins were first purified by metal affinity using a HisTrap HP 5 ml column (GE Healthcare) equilibrated with buffer A of DEN2 FLNS5 without β-ME (Table 2.3). After lysis, all supernatants were flowed through the HisTrap HP 5 ml column. Unbound proteins were washed away sequentially with five column volumes buffer A supplemented with 25 mM and 125 mM imidazole respectively. Proteins were then eluted using a linear gradient of imidazole from 125 to 500 mM. Fractions that contain the proteins were concentrated and subjected to gel filtration purification. As a final polishing step protein were further purified with HiPrep 16/26 Superdex 75 in buffer containing 20 mM Tris HCl pH 7.0, 150 mM NaCl, 10% Glycerol and 1 mM EDTA. The eluted monomeric proteins were concentrated to 20 mg/ml and used for crystallization assays or functional studies.

2.5 Protein refolding

Recombinant NS5 proteins spanning NS5 residues 341-900 and 415-900 expressed from pET 32a+ (protein 1 and 5, Table 2.2) were unfolded overnight in 20 mM Tris HCl pH 7.5 and 10

mM β-Mercaptoethanol (Buffer D) supplemented with 6 M GdnHCl. The proteins were subsequently refolded by column refolding method, with the chaotrope slowly removed. Briefly, the unfolded proteins loaded in the HisTrap HP column (GE Healthcare) were washed with Buffer D supplemented with 6 M Urea, 0.5 M NaCl and 20 mM imidazole. Proteins were then refolded with Buffer D of DENV 2 FLNS5 that was supplemented with 0.5 M NaCl and 20 mM imidazole and eluted with buffer D supplemented with 0.5 M imidazole.

2.6 NS5 RdRp activity measured using Scintillation Proximity Assay (SPA)

The NS5 RdRp assays were performed in white half-area flat-bottom 96-well plates (Corning) in a buffer containing 50 mM Tris-HCl, 10 mM KCl, 2 mM $MgCl_2$, 2 mM $MnCl_2$ at pH 7.0, with 0.25 μg poly C annealed to 0.0125 μg biotinylated oligo G20 (Sigma) as a template, and 1.75 μg of DENV 3 RdRp protein in a volume of 22.5 μl. Following incubation for 1 h, 2.5 μl of 4 μM of total GTP containing 0.5 μCi of [^{3}H]GTP (6.1 Ci/mmol) (GE healthcare) was added to initiate the reaction. For measuring the IC50 of the 3'dGTP inhibitor, the enzyme was pre-incubated with template and inhibitor for 1 h before the reaction was started with addition of solution containing radioactive GTP for 1 h. The reaction was stopped with EDTA containing streptavidin coated SPA beads (GE healthcare). All reactions were carried out at room temperature.

2.7 Limited proteolysis

DENV 2 full length NS5 was incubated with trypsin (Sigma), in a ratio of trypsin/ enzyme (20:1 or 1: 100) in 20 mM Tris HCl pH 7.5, 50 mM NaCl, 10 mM β- Mercaptoethanol and 2 mM $CaCl_2$ at room temperature or 37°C. The reactions were quenched with SDS-containing sample buffer, and analyzed by SDS-PAGE. The desired bands in SDS PAGE gel were

transferred to a polyvinlidene fluoride (PVDF) membrane using a Trans-Blot electrophoresis transfer cell (BioRad) for 1 h at 4°C. N-terminal amino acid sequencing was performed either at the Novartis Institute for Biomedical Research in Basel or Protein and Proteomics Centre of National University of Singapore. A similar procedure was performed for DENV polymerase catalytic domain (DENV 1-3) with trypsin/enzyme ratio at 1: 1000 at room temperature for 30 min.

2.8 Fab antibody biopanning

A fully human Fab fragment with a library size of 1.2 $x10^{10}$ (Morphosis) was used for biopanning to select antibodies specficially against DENV 2 FLNS5 for antibody selection. In this case, a serotype of DENV 2 TSVO1 NS5 was used. The biopannning was conducted in collaboration with Dr Celine Nkenfou (NITD), at the Novartis Insitute for Biomedical Research in Basel (details refer to Morphosis protocol, internal use only). Prior to that, an experimental course in antibody phage was provided for this work (attended by Dr. Celine Nkenfou and myself) to master the biopanning method (in Switzerland). Briefly, DENV 2 FLNS5 was diluted to 50 µg/ml and 300 µl of that coated to a maxisorp microtiter plate (Nunc) for overnight at 4°C. The MBP-His proteins were used as control. The solution in the coated plate was washed 2 times with PBS. Subsequently, the wells were blocked with 400 µl 5% MPBS for 2 h at RT on a microtiter shaker. The phage solution from the library was mixed with 2x blocking solution (10% MPBS) and incubated for 2 h at RT to reduce the binding of phage to milk proteins. After the blocking procedure, the wells were washed 2 times with 400 µl PBS. 300 µl of the pre-blocked phage was mixed into blocked well and incubated for 2 h at RT. The unspecific phage was washed with 400 µl PBS for 3 times. The specifically bound phages were eluted by adding 300 µl of 20 mM DTT in 10 mM Tris HCl, pH 8.0 into each well and was incubated for 10 min at RT without shaking. The eluted

solution was added to *E.coli* TG1 that was pre-incubated at 37°C for 45 min in an incubator without shaking and tirated as plaque forming units (pfu). The selected phage was rescued as decribed in the protocol and the panning process was repeated for a further 2 selection rounds. After the panning selection, the selected Fab fragments were subcloned into expression vector. The Fab antibody binding to the antigen was validated by 2 rounds of ELISA. Five positive clones from the secondary ELISA screening were confirmed by sequencing.

2.9 ELISA assays

Binding of monoclonal antibody Fab fragment was determined using ELISA assays. Briefly, 5 µg/ well of proteins were coated overnight at 4° C in 96-well microtiter plates (Nunc). After blocking with 5% MPBST for 1 h at RT, wells were washed with 3x PBST. Appropriate dilutions of antibody (1:5000 dilutions) were then added to the wells and incubated for 1h at RT. The wells were washed with 3x PBST. Bound antibody were detected using goat anti-Human IgG F(ab') 2 alkaline phosphatase (Jackson Immuno) and the colorimetric substrate-p-nitrophenyl phosphate. A_{405} meassurement (absorbance at 405 nm) was followed with time using a plate reader (Molecular Devices), with values corrected by subtracting both the absorbance at 0 min and that for wells incubated without proteins. To deterimine the K_d for respective Fab fragments against NS5, the procedure was as follow. The purified antibodies were serially diluted into concentration of 10 µM, 1 µM, 100 nM, 10 nM, 1 nM and 0.1 nM and added to coated NS5 on ELISA plate for 1 h at RT. The subsequent steps for detection and measurement were similar as above.

2.10 Western blot analysis

Proteins were separated on 10-12 % SDS PAGE gel and transferred onto a nitrocellulose membrane either using Semi-Dry (Bio-Rad) to confirm the presence of desired protein.

Briefly, NS5 related proteins were incubated with primary mouse anti-His antibody at 1:2500 dilution or anti-NS5 polyclonal antibody at 1:5000 dilution for 1 h, followed by secondary peroxidase-conjugated goat anti mouse IgG at 1:2500. Detection was carried out as described in the manufacturer's manual.

2.11 Flame analysis

An analysis of metals bound to the DENV-3 polymerase (purified as described above, in 86 μM) was diluted to a final concentration of 5 μM in MiliQ water using flame atomic absorption spectroscopy with a polarized Zeeman Z-2000 atomic absorption spectrophotometer (Hitachi). Solution that contains $MgCl_2$, $ZnCl_2$ and $NiCl_2$ were used as respective reference.

2.12 Crystallization

Crystals of DENV 3 RdRp were obtained using the hanging drop vapor diffusion method at 4°C. Typically, a volume of 0.75 μl of protein solution at a concentration of 11 mg/ml in 20 mM Tris HCl pH 6.8, 0.25 M NaCl, 1mM EDTA, 2 mM β-Mercapto-ethanol, 0.1% w/v CHAPS was mixed with an equal volume of a reservoir solution (0.1 M Tris HCl at pH 8.5, 0.8 M K/Na tartrate, 0.5% w/v PEG MME 5k). Lens-shaped crystals with a blueish colour belonging to space group $C222_1$ with unit-cell parameters of a= 160.28 Å, b= 178.77 Å, and c= 58.05 Å were obtained after three to four weeks, but they diffracted very poorly.

A multi-step strategy was subsequently employed in order to improve crystal diffraction (described in chapter 4). Briefly, crystal quality was markedly improved by the addition of $MgSO_4$ or Mg acetate at concentrations of 0.005-0.2 M and by subjecting crystals to a careful dehydration protocol. Prior to data collection, crystals were dehydrated in a stepwise manner

by soaking them into a series of buffers containing mother liquor (5 µl) with a gradual increase of PEG 3350 concentration, ranging from 15%, 25% to 35% (w/v). Crystals were then mounted in a cryo-loop and rapidly cooled to 100 K in liquid nitrogen.

2.13 Data collection

The diffraction intensities for PDB 2J7U and 2J7W were recorded using 1° oscillations on CCD detectors (ADSC and MAR225) on beamline ID14-4 at the ESRF (Grenoble, France) or X10SA at the Swiss Light Source (Villigen, Switzerland) using attenuated beams. Evaluation of crystal quality was performed with an R-AXIS IV^{++} imaging-plate detector mounted on a Rigaku/MSC FR-E X-ray generator. Integration, scaling and merging of the intensities were carried out using the programs MOSFLM (Leslie, 1999) and SCALA from the CCP4 suite (Bailey, 1994).

2.13 Soaking experiment

(1) The RdRp-Mg^{2+} complex structure was obtained by soaking crystals in 90 mM Tris HCl at pH 8.5, 0.2 M $MgCl_2$ and 25% w/v PEG 3350 for 20 h at 4 ºC. In the case of RdRp-Mn^{2+} complex, crystals were soaked in the similar condition as in Mg^{2+} complex but with lower concentration to 10mM $MnCl_2$. See chapter 4 for detail. (2) To obtain a complex with 3'-deoxy-guanosine-5'-triphosphate (3'dGTP), DENV 3 RdRp crystals were soaked in 3'dGTP (Trilink) in the mother liquor supplemented with 10 mM 3'dGTP for 20 h at 4 ºC. Crystals were then dehydrated using a similar protocol as described above, before being flash-cooled in liquid nitrogen. (3) X-ray crystallography inhibitors screening were also conducted with novel compounds synthesized as part of a lead optimization (lead identified from a high throughput screening campaign using DENV 2 FLNS5) program at NITD. Inhibitors were

provided by Dr Yin Zheng and Dr Chen Yen Liang (NITD). Detail soaking procedure for ligands/substrates is discussed in chapter 4.

2.14 Structure solution, refinement and analysis

(1) The structure of PDB 2J7U and 2J7W were solved by molecular replacement using the WNV polymerase domain structure as an initial model (PDB code 2HFZ) (Malet et al., 2007). The refinement was carried out using programs CNS (with molecular dynamics in torsion angles space) and REFMAC5 with the TLS refinement option. The latter proved important to model several mobile regions of the fingers subdomain. Refinement cycles were interspersed with model rebuilding sessions at an SGI computer graphics with program O. The coordinates have been submitted to the PDB with accession numbers 2J7U and 2J7W for the DENV 3 NS5 Mg^{2+} and 3'dGTP complex structures respectively. Structure related figures were drawn using program PyMOL. Superpositions of structures and calculations of the r.m.s deviations were carried out using program LSQKAB from the CCP4 suite. (2) To identify ligands (inhibitors, RNA template and Mn^{2+}), program AMoRe was used for molecular replacement with the unliganded DENV 3 RdRp structure (PDB code 2J7U) as a search model. Refinement was carried out with the REFMAC5.

CHAPTER 3

PROTEIN EXPRESSION STRATEGY

AND

CHARACTERIZATION OF NS5 POLYMERASE

INTRODUCTION

The success of genomic sequencing projects in recent years offers opportunities to characterize a large percentage of identified gene products with unknown functions. Today, X-ray crystallography and NMR technique have become the main tools to address the needs in postgenomic era and are an integral part of drug discovery programs (Murphy, 1996). Proteins suitable for structural characterization often require having properties such as good expression yield, solubility, purity, monodispersity and stability. Expressions of soluble proteins have always been one of the major bottlenecks in the field. For example, many biologically interesting mammalian (human) proteins or membrane associated proteins are difficult to express as a soluble protein using bacterial expression systems. As a result, there has been a decline in the reported number of soluble heterologous proteins expressed in *E.coli* (Chatterjee and Esposito, 2006). Several strategies have been adopted to improve the situation by using chaperones and also engineered strains to provide a "eukaryotic cell like" environment in *E.coli* (Baneyx and Mujacic, 2004).

Many advancement have been integrated into the structure determination pipeline. These comprise advances made in areas of (1) computational biology and bioinformatics (Goldsmith-Fischman and Honig, 2003), (2) protein engineering (Derewenda, 2004), (3)

protein expression and purification (Goulding and Perry, 2003) and (4) methods for accelerating crystallization and X-ray data collection (Leslie et al., 2002).

The Flaviviridae polymerases from DENV, HCV and BVDV have been shown to be difficult proteins for structural studies due to solubility and crystallizability issues. The first active DENV polymerase from DENV 1 was expressed in *E.coli* by Tan et al (Tan et al., 1996). Subsequently, Ackermann et al., 2001 described the expression of an active DENV 2 FL NS5 in *E.coli* that contains polymerase activity (Ackermann and Padmanabhan, 2001). Nevertheless, the yield and quality of these purified proteins were insufficient for high throughput screening assays (HTS) or structural studies. In the case of HCV success in crystallization was achieved when the last 21-55 hydrophobic amino acids at C-terminus were truncated. These proteins were also used in HTS screening (Wang et al., 2003). In the case of BVDV polymerase, after the initial protein construct (residues 1-695) failed to produce protein crystal, a combination strategy of limited proteolysis, homology modeling and high throughput cloning and expression was employed. A total of 84 protein constructs were designed and only two of these proteins produced crystal structures (Choi et al., 2004).

Objective of the study

The aim of the study reported in this chapter was to explore strategies that can lead to soluble DENV NS5 protein expression. These include (1) expression systems using the insect or bacterial cell hosts, lysis buffer and expression temperature optimization, fusion tag and host-vector relationship. (2) refolding, (3) limited proteolysis and (4) homologous protein screening (Fig 3.1). All NS5 related proteins that derived from these strategies were used for functional and structural studies. In the second part of this study, we aim to characterize NS5 polymerase using several approaches for understanding its roles and function.

RESULTS AND DISCUSSION

3.1 Overview of protein expression strategy

An active and stable DENV 2 FL NS5 (TSV01, 1-900 amino acid) was successfully expressed in a baculovirus system with pBakpak 9 vector (Fig 3.1 & 3.2a) and the purified protein was used for crystallization screening .

Column refolding was applied on NS5 domain proteins (starting from residue 341-900 and 415-900), after various unsuccessful attempts to solubilize these proteins. Whilst not useful for structural studies, these proteins were useful for antibody mapping (see section 3.3.2 & Fig 3.1 & 3.2b).

Limited proteolysis was therefore employed to generate protein variants that would be suitable for crystallization assays (protein 12-18, Table 3.1, Fig 3.1& 3.2c). Limited proteolysis of the DENV 2 FLNS5 protein alloewed us to identify stable fragments starting at residues 48, 264, and 359. The soluble proteins generated from this approach are 48-359, 359-866 and 264-359 (residue number) (protein 12-13 & 18, Table 1, Fig 1& 2c). Some of these proteins were also used in antibody mapping (see later).

Homologous protein screening was useful to deduce soluble, stable and active DENV or WNV polymerases. The following homologous protein were used to screen for protein solubility - (1) RdRp domain of DENV 1-4, (2) DENV 2 FLNS5 (NGC strain) and (3) WNV FLNS5 (Fig 3.1). Of the four NS5 RdRp constructs the best results were obtained with DENV 3 RdRp truncation construct staring at amino acid 273. In total, we have designed and expressed 25 NS5 related protein constructs for functional and structural studies (Table 3.1, 3.2).

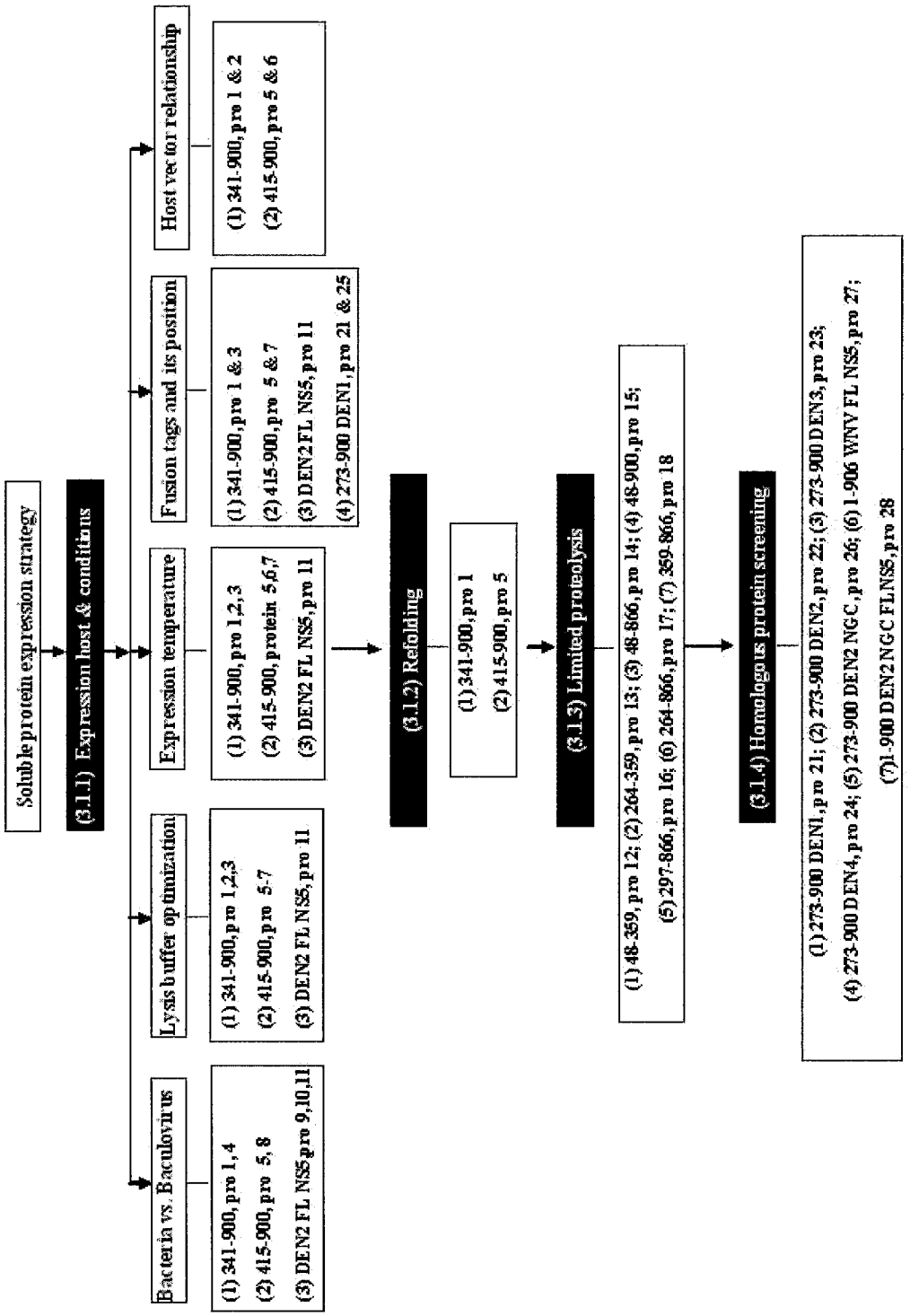


Fig 3.1.Flowchart for strategies in identifying protein designs.

The approach involves (1) use of expression hosts and conditions which comprises bacteria vs. baculovirus system, lysis buffer optimization, expression condition and fusion tags (3.1.1). Active and soluble DENV 2 FLNS5 protein were expressed in insect cells while domain protein 341-900 or 415-900 were expressed insoluble. (2) Refolding of the truncated domain proteins (3.1.2). (3) Use of limited proteolysis to generate 3 soluble proteins from 7 constructs (3.1.3). (4) Use of homologous protein screening and identified 5 soluble proteins out of 7 constructs (3.1.4). See texts for details. Proteins that were designed with various respective approaches were listed numbered as shown in Table 3.1. Solubility and yield for respective proteins are also shown in Table 3.1.

Table 3.1

No	**Protein (amino acid)**	**Plasmid**	**Strategy used to solubilize protein**	**Expression host**	**Solubility**	**Yield (mg)/ L**
1.	341-900 DENV2	pET 32a	-Detergents -Fusion tag (Trx) -Refolding	BL 21 (DE3) RIL	1-5%	1 mg after refolding
2.	341-900 DENV2	pET 32a	-Fusion tag (Trx) -host cell	Origami (DE3)	1-5%	N.A
3.	341-900 DENV2	pET 21b	-Fusion tag (His) position	BL 21 (DE3) RIL	Not soluble	N.A
4.	341-900 DENV2	pFastBac HTB	-Baculovirus expression	High five cells	5- 10%	0.4-0.5
5.	415-900 DENV2	pET 32a	-Detergents -Fusion tag -Refolding	BL 21 (DE3) RIL	Not soluble	1 mg after refolding
6.	415-900 DENV2	pET 32a	-Fusion tag (Trx) -host cell	Origami (DE3)	Not soluble	N.A
7.	415-900 DENV2	pET 21b	-Fusion tag (Trx) position	BL 21 (DE3) RIL	Not soluble	N.A
8.	415-900 DENV2	pFastBac HTB	-Baculovirus expression	High five cells	Not soluble	N.A

Continue Table 3.1

9.	1-900 DENV2	pFastBac HTB	-Baculovirus expression	High five cells	70-80%	4-5
10.	1-900 DENV2	pBakPak 9	-Baculovirus expression	High five cells	70-80%	8-10
11.	1-900 DENV2	pET 32a	-Fusion tag (Trx)	BL 21 (DE3) RIL	70-80%	3-5
12.	48-359 DENV2	pGEX 4T1	-Limited Proteolysis	BL 21 (DE3) RIL	60-70%	2-3
13.	264-359 DENV2	pGEX 4T1	-Limited Proteolysis	BL 21 (DE3) RIL	60-70%	3-4
14.	48-866 DENV2	pET 32a	-Limited Proteolysis	BL 21 (DE3) RIL	1-2%	N.A
15.	48-900 DENV2	pET 32a	-Limited Proteolysis	BL 21 (DE3) RIL	1-2%	N.A
16.	297-866 DENV2	pGEX 4T1	-Limited Proteolysis	BL 21 (DE3) RIL	Not soluble	N.A
17.	264-866 DENV2	pET 32a	-Limited Proteolysis	BL 21 (DE3) RIL	Not soluble	N.A
18.	359-866 DENV2	pET 32a	-Limited Proteolysis	BL 21 (DE3) RIL	20-40%	3-4
19.	1-296 DENV2	pQE 30		M15 or SG	30-50%	1-2
20.	1-296 DENV2 NGC	pQE 30		M15 or SG	30-50%	1-2
21.	273-900 DENV1	pET 15b	-Homologous Protein Screening	BL 21 (DE3) RIL	10-20%	0.2-0.3
22.	273-900 DENV2	pET 15b	-Homologous Protein Screening	BL 21 (DE3) RIL	15-25%	0.4-0.5
23.	273-900 DENV3	pET 15b	-Homologous Protein Screening	BL 21 (DE3) RIL	60-80%	3-5
24.	273-900 DENV4	pET 15b	-Homologous Protein Screening	BL 21 (DE3) RIL	5-15%	0.1-0.3
25.	273-900 DENV1	pDEST 14	-Fusion tag position	BL 21 (DE3) RIL	60-80%	1-3
26.	273-900 DENV2 NGC	pET 28a	-Homologous Protein Screening	BL 21 (DE3) RIL	30-40%	1-2

Continue Table 3.1

27.	1-906 WNV	pET 28a	-Homologous Protein Screening	BL 21 (DE3) RIL	60-80%	3-5
28.	1-900 DENV2 NGC	pET28a	-Homologous Protein Screening	BL 21 (DE3) RIL	50-60%	1

Table 3.1.Summary of expression.

List of expression host, vector, solubility profile and yield for all NS5 related proteins used in this study. For DEN2 NS5 serotype, there is 2 viral strains used in this study- of that is TSVO1 and NGC strain. DEN2 in this table is refered to as DEN2 TSVO1.

Table 3.2

Protein used for	**Proteins (remark)**
Elisa Assay & immunoblotting for antibody epitope mapping	341-900 DENV2 (pro 1) 415-900 DENV2 (pro 5) 48-359 DENV2 (pro 12)# 264-359 DENV2 (pro 13)# 359-866 DENV2 (pro18)# 273-900 DENV1 (pro 25) 273-900 DENV2 (pro 22) 273-900 DENV3 (pro 23) WNV FLNS5 (pro 27) DENV 2 FLNS5 (pro 10) DENV 2 FLNS5 NGC(pro 28)
Crystallization assays	273-900 DENV1 (pro 25) 273-900 DENV3 (pro 23) WNV FLNS5 (pro 27) DENV 2 FLNS5 (pro 10) DENV 2 FLNS5 NGC (pro 28) MTase DENV2
Enzymatic assays	WNV FLNS5 (pro 27) DENV 2 FLNS5 (pro 10) DENV 2 FLNS5 NGC (pro 28) 273-900 DENV1 (pro 25) 273-900 DENV2 NGC (pro 26) 273-900 DENV3 (pro 23)*

Continue Table 3.2

Protein used for	**Proteins (remark)**
Limited proteolysis	DENV 2 FLNS5 (pro 10) 273-900 DENV1 (pro 25) 273-900 DENV2 NGC (pro 26) 273-900 DENV3 (pro 23)

Table 3.2.Summary for NS5 related proteins used for functional and structural characterization. # Large scale expression and purification were attempted for crystallization assays, but aborted due to proteolytic problem. * Enzymatic assays result shown in this report. Protein number is listed as in Table 3.1.

3.1.1 Expression hosts and conditions

Recombinant protein expression of DENV 2 NS5 (TSVO1 strain) - (1) residues 341-900, (2) 415-900 and (3) FL NS5 were investigated using baculovirus and bacterial host expression system. Based on the hypothesis that the NLS region of NS5 could form a separate flexible domain, constructs starting at positions 341 (341-900) and 415 (415-900) were designed. The bacterial or baculovirus derived proteins were expressed insolubly, even after initial optimizations in combination with lysis buffer optimization (detergent, salt, pH), expression conditions and fusion tags (Fig 3.1).

Fusion tags- hexahistidine (His), maltose binding protein (MBP), glutathione-S-transferase (GST) and thioredoxin (Trx) are reported to often enhances solubility of proteins (Chatterjee and Esposito, 2006). His, Trx and GST fusion tags were incorporated into protein design (Table 3.1). Only the change of His fusion tag position remarkably affected the solubility of DENV 1 RdRp (273-900). The C-terminal His-tagged DENV 1 RdRp yielded more soluble protein compared to those of N-terminal His tag (protein 21 vs. 25 of Table 3.1). Protein amino acid sequence determines protein stability, isoelectric point (pI), hydrophobicity and

molecular weight, all of which contribute to protein expression and solubility (Esposito and Chatterjee, 2006). So, fusion tags could change the solubility of target protein with its amino acids sequence, particularly the bigger fusion tags such as Trx or GST tags. Some successful cases of crystallization with large GST or Trx fusion tags have also been reported (Smyth et al., 2003).

The lack of posttranslational modifications or eukaryotic chaperones could lead to accumulation of misfolded protein, forming protein aggregates in the cytoplasm known as inclusion bodies, which are possible reasons that suggest why some NS5 contructs resulted in insoluble form. Despite these drawbacks, *E.coli* cells remain as the major expression host for structural studies because it is easy to use, cheap, and quick to make target gene modifications. For example, site directed mutagenesis was applied on DENV 3 RdRp to improve the protein stability for structural studies (see appendix A2 for description).

After unsuccessful soluble or stable protein expression in *E.coli* cells, eukaryotic cell expression was attempted using the baculovirus expression. DENV 2 FLNS5 (TSV01 strain) expressed in Hi5 insect cells for 72 h (pBakpak 9 vector, protein 10, Table 3.1) shows relatively good protein yields and stability. Each liter of culture produces about 20 mg protein (Fig 3.2a, lane 3), which was used for crystallization (> 1000 screening conditions) and functional studies (eg. to test for enzymatic assay, see 3.3.3). Insect cells expression allows proper heterologous protein folding in cytoplasm, post-translational modification or oligomerization in manners similar to those of mammalian cells (Kost et al., 2005). In contrast, DENV 2 FLNS5 protein (TSVO1 strain) produced in *E.coli* cells (protein 11, Table 3.1 & Fig 3.2a) was strongly susceptible to proteolysis and severely degraded throughout the purification process. In the case of truncated RdRp protein (341-900 and 415-900), neither

E.coli nor baculovirus system produced protein yielded soluble protein for structural studies (protein 1-8, Table 3.1).

3.1.2 Refolding

Insoluble truncated NS5 proteins 341-900 and 415-900 were refolded using chromatographic approach. The recovery yield for both proteins was about 0.8-1 mg proteins/L. However, the C-terminus of proteins were sensitive to degradation (Fig 3.2b) and only used for ELISA assays and antibody epitope mapping (Table 3.2). There is no universal method of refolding proteins so far. In principle, inclusion bodies can be solubilized through the use of high concentration of denaturants such as urea or guanidine hydrochloride, assisted by the addition of reducing agent (e.g. 1-500 mM DTT or β-ME). The solubilized protein can then be refolded by slow removal of the denaturants either by dilution (dialysis) or refolding column (Singh and Panda, 2005).

3.1.3 Limited proteolysis strategy

The initial attempts to crystallize the DENV 2 FLNS5 (TSVO1 strain) that produced in baculovirus system for crystallization screenings (>1000 conditions) were unsuccessful. This prompted us to use limited proteolysis strategy to identify stable protein constructs which may lead to successful crystallization. Stable fragments were identified by N-terminal sequencing, after DENV 2 FLNS5 protein was digested with trypsin to cause limited proteolysis (Fig 3.2e, lane 2). Based on the results of tryptic digest, 7 truncation constructs (starting at residue 48, 264 and 359) were designed (protein 12-18, Table 3.1). All of minigenes encoding the various constructs share universal cloning sites (Bam HI and Sal I), in order to simplify the PCR amplification, restriction digestion and ligation using either pGEX 4T1 or pET32a expression vector. Three soluble proteins were expressed, namely (1)

GST tagged 48-359, (2) GST tagged 264-359 and (3) Trx tagged 359-866 (Fig 3.2c) (indicating as the start-end aa of NS5).

The GST tagged proteins contain part of the NLS region, which has been known as the bNLS (eg. Importin-b binding site). Structural insight into these proteins might provide clues to its function and role. However, these proteins were susceptible to proteolysis (Fig 3.2c, lane 1,2 & 4). The Trx tagged 359-866, also displayed proteolysis at the C-terminal. Enzymatic assay shows that the protein (359-866) is inactive (data not shown) and is consistent with the study of Malet. H et al., 2007 and with that our study in elucidating DENV 3 RdRp structure (see later). From the solved structure, the lack of activity of the protein can be attributed to finding that NLS forms a component of the polymerase fingers subdomain and therefore is important for polymerase activity (see chapter 5). Nevertheless, these GST or Trx tagged proteins were useful in ELISA assays and antibody epitope mapping (Table 3.2).

DENV polymerase has only less than 20% sequence identity to those of BVDV and HCV polymerases for which structures exist in the PDB database. The low sequence identity means that homology modeling was not applicable in designing DENV NS5. Limited proteolysis is an useful approach when one encounters problem to get soluble or stable native protein or difficulties in obtaining protein crystals (Christ and Winter, 2006; Vasilyeva et al., 2004). Many viral proteins such as DENV NS5 contain multiple domains to perform different functions within the same polypeptide chain. Due to evolutionary selection, domain boundaries prediction by computational methods can be obscured by mutations, insertions and deletions near the boundary. Therefore the actual domain boundary of a protein may be different from the prediction (Staub et al., 2004). The rationale of using limited proteolysis is to make use of protease cleavage sites which tend to be located in the flexible regions of the

polypeptide chain and less compact regions (Novotny and Bruccoleri, 1987), which would further lead to the identification of protein domain. The exposed sites have often been shown to have a good correlation with large crystallographic *B*-factors and poorly defined electron density, as evidenced from Thermolysin study (Fontana et al., 1986). In fact, the NLS region that lies between MTase and RdRp is often more susceptible to proteolysis. This is consistent with the tryptic digest constructs at residues 264 and 359 and MTase DENV 2 NS5NGC (1-296 aa) crystal structure. A high *B*-factor is shown within residues 261-264 of MTase and no electron density was observed after this region (Egloff et al., 2002).

3.1.4 Homologous protein screening

We also sought to use homologous protein screening to design constructs that could be expressed in *E.coli* in order to increase the likelihood of obtaining crystals. This approach comprises the catalytic domain protein (RdRp) and full length protein screening (FL). After initial screening of protein expression for NS5 RdRps with the four DENV serotypes, the best results were obtained with one DENV 3 RdRp truncation construct starting at amino acid 273 (Fig 3.2d, lane 5; protein 23 Table 3.1). Taken together with the constructs that were designed and expression results discussed before (3.1.1-3.1.3), suggests that inclusion of the NLS region was essential to obtain a stable protein. In addtion, the C-terminal His tagged DENV 1 (273-900) also has soluble protein fraction (Fig 3.2d, lane 6; protein 25 vs. protein 21 in Table 3.1). Both the DENV 1 and DENV 3 RdRp catalytic proteins could be expressed in high quantity (about 1-3 mg per liter of culture) and were purified to >95% homogeneity. These proteins were subjected to further functional and crystallization studies. The catalytic domains from other serotypes (DENV 2 & 4, protein 21 & 24 Table 3.1) were discontinued because of low protein expression levels and solubility problems. Interestingly, higher solubility of DENV 2 RdRp (NGC strain) (Fig 3.2d, lane 8; protein 26 Table 3.1) was found

compared to the same serotype of DENV 2 (TSVO1 strain) (Fig 3.2d, lane 7; protein 22 Table 3.1).

The FLNS5 has two useful proteins derived from homologous protein screening. A good yield and stable N-terminal His tagged WNV FL NS5 was expressed in *E.coli* (3-5 mg/L culture), compared to DENV FLNS5 (Fig 3.2a & 3.2d). The NLS regions for WNV NS5 was better in resisting proteolysis, yielding only MTase fragment and FL proteins (Fig 3.2d, lane 1; protein 27 Table 3.1). The DENV 2 FLNS5 (NGC strain) has also similar proteolytic profile as WNV NS5. However, the expression yield is lower (<1 mg/L culture). In both cases the cleaved MTase was easily removed by gel filtration and the FL protein was concentrated for crystallization assays (Fig 3.2d). In the case NS5 homologous screening, this approach has learnt that cloning different strains or serotypes representing naturally occurring variants of a target protein from an infectious agent (e.g. from a bacterium or a virus in this case), greatly increases the likelihood of obtaining a protein sample amenable to crystallization, as compared to a single serotype or variant, as demonstrated by other workers in the field of structural genomics (Au et al., 2006).

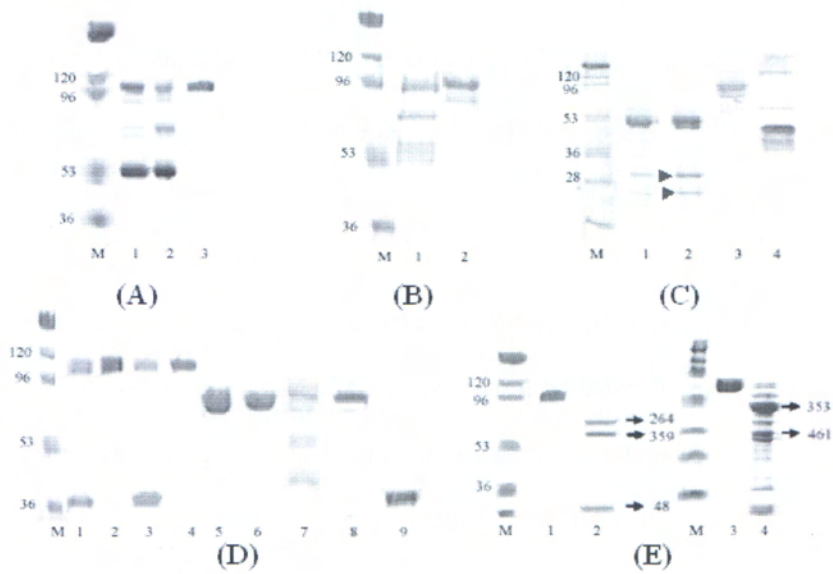


Fig 3.2.SDS PAGE gel analysis.

NS5 related proteins expressed and purified for structural and functional studies in this report. In total, 25 NS5 related proteins were designed, expressed and screened, 13 of those were sufficiently expressed and used for functional and structural studies and only 6 of those fulfilled the prerequisite crystallization assays condition. M is denoted as marker. (A) Profile of DENV 2 FL NS5 (TSVO1 strain). Lane 1 and 2, bacterial expressed DENV 2 FLNS5 after metal affinity purification and ion exchange purification respectively. Protein was degraded along the purification. Lane 3, baculovirus expressed stable DENV 2 FLNS5 after metal affinity column. (B) SDS PAGE gel profile of protein 415-900 and 341-900 NS5 (TSVO1 strain) that refolded by chromatographic method (Lane 1 and 2 respectively). (C) Profile of soluble NS5 proteins (TSVO1 strain) derived from limited proteolysis analysis. Lane 1 and 2, GST tagged 48-359. Protein degraded along the purification (lane 2). Two smaller size products, indicated by arrow (near 28kDa and below) are GST fusion tag and native 48-359. Lane 3, Trx tagged 359-866. Lane 4, protein GST tagged 264-359. (D) Profile of soluble protein deduced from homologous protein screening approach. Lane 1, WNV FLNS5 after metal affinity purification. Degraded MTase was observed. Further gel filtration purification removed MTase and retained the WNV FLNS5

protein (Lane 2). A similar profile of FLNS5 protein was screened for DENV 2 FLNS5 (NGC strain) (lane 3 and 4). Lane 5-8 represents DENV3, DENV1 and DENV2 (TSVO1 strain) and DENV2 (NGC strain) RdRp respectively. The best protein yield was that of DENV3 and DENV1 RdRp. DENV 2 RdRp (TSVO1 strain) is less stable than other serotype RdRp. Lane 9, MTase DENV 2 1-296 that expressed for structural studies. (E) Limited proteolysis of the DENV 2 FL NS5 protein (left) and of the DENV 3 RdRp catalytic domain (right). Lane 1 and 2, undigested FL NS5 and trypsin-digested NS5 respectively. The N-termini of digested protein fragments were sequenced (264, 359 and 48), and the information was used to derive constructs shown in table 1. Lane 3 and 4, undigested DENV 3 RdRp and trypsin-digested DENV 3 RdRp, respectively. A similar tryptic digest profile was found for DENV 1 and DENV 2 NGC RdRp. The trypsin/ enzyme ratio for FL NS5 was 1:100 and the reaction was carried out at room temperature for 2 h (lane 3). Using a lower concentration of trypsin (1:1000 ratio of trypsin to RdRp) for ~ 30 min at room temperature yielded several smaller fragments (lane 4). The N termini of digested protein fragments were sequenced (indicated by arrow).

3.2 Characterization of NS5 polymerase

3.2.1 Enzymatic Activity

In this study, the enzymatic activity of DENV 1-3 RdRp catalytic domain and DENV 2 FLNS5 were conducted. The dengue unit lab at NITD has several DENV2 viral strains. Both the NS5 TSVO1 and NGC strain were expressed as DENV FLNS5 protein. The former was expressed in baculovirus and the latter in *E.coli* cells. The FLNS5 (TSVO1 strain) protein was first constructed (A Brooks, James Cook University). It has been used in HTS campaign and initial crystallization efforts (see section 3.3). The FLNS5 NGC strain constructed using *E.coli* as expression system later was used in Fab co-crystallization (see later). In addition, the NGC strain RdRp catalytic domain protein was also used in enzymatic activity studies.

This was because the TSVO1 strain RdRp was not as soluble as the NGC strain RdRp. Similarly, DENV 4 RdRp was not a soluble protein. Therefore the comparison of catalytic domain (RdRp) activity was tested with RdRp of DENV1, DENV2 NGC and DENV3.

3.2.1.1 Assay for DENV 3 RdRp

A scintillation proximity assay (SPA) was set up using homopolymer C as template and biotin-oligoG20 as primer to determine the activity of recombinant proteins. The set up of this RdRp assay was based on the optimized assay condition of DENV 2 FLNS5 (Niyomrattanakit et al., manuscript in preparation, see publication). The purpose of the assay was to evaluate whether the protein that crystallized was an active enzyme. In this assay, the addition of tritiated GTP to a primer is measured directly without the need for any wash steps since only the radioactive decay of tritium on the primer/template complex that is in close proximity to the scintillant via streptavidin/biotin interaction will result in light emission that can be measured. Figure 3.3a shows the activity of the truncated DENV 3 RdRp (residue 273-900), monitored over time. Approximately to 1 μM DENV 3 RdRp was used to achieve signals that are comparable to those for DENV 2 FLNS5 (50 nM; data not shown), implying that the domain is ~ 20 fold less active than the full-length protein. Nevertheless, the sensitivity, speed and possibility to miniaturize the assay to a 384-well format will make the assay valuable for a high-throughput screening of inhibitors. The ability of 3'deoxyguanosine triphosphate (3'dGTP) to block the elongation process was studied using the NS5 RdRp-SPA format. The 3'dGTP is a chain terminator, lacking a 3'OH group that is required for the formation of phosphodiester bond between two nucleotides during replication. Therefore, incorporation of 3'dGTP into the nascent strand will terminate the daughter strand extension. In this experiment, the enzyme, template/primer complex and the inhibitor at various concentrations were incubated for 1 h prior to initiation as above. The data presented in Fig

3.3b show that the IC_{50} of the DENV 3 RdRp is 0.22 μM (compared with an IC_{50} of 5 nM for FL NS5: data not shown), and demonstrate that the 3'dGTP can serve as a useful reference compound in the present assay format for finding new inhibitors.

3.2.1.2 Comparison of different serotype DENV RdRp and DENV FLNS5

As discussed earlier, the DENV 3 RdRp is ~ 20 fold less active than the full length protein. Unexpectedly, we found that the DENV 2 RdRp (NGC strain) activity was ~ 4 folds less active than DENV 2 FLNS5, under similar experiment conditions (data not shown). We observed that the ezymatic acitivity (active to less active, in chronological order) as following - FL NS5 >DENV 2 RdRp (NGC) >DENV1 or DENV 3 RdRp, respectively. First, the catalytic domain of DENV1, DENV2 and DENV 3 RdRp (residue 273-900) possibly adopt a more open conformation than that of DENV 2 FLNS5, which causes the domain proteins to be less active. Limited proteolysis resulted in more protease accessible sites in the DENV RdRp catalytic domain (using DENV1, DENV2 NGC and DENV3) than in the FL (Fig 3.2e, lane 4), since similar amounts of trypsin was used. Limited proteolysis is useful to probe conformational changes of proteins by pinpointing the sites of local unfolding in a protein chain (Huang, 2003; Hubbard, 1998). In the case of DENV 2 RdRp (NGC strain), the activity of this protein is higher than other serotypes (DENV1 & DENV3). Variation at the amino acid sequence level among the DENV 1-3 RdRp serotypes may accounts for the varying activity of serotypes and could also impact the protein folding. These suggestions have to be validated by NMR experiments. Interestingly, when DENV1 and DENV 3 RdRp proteins stock were diluted into a new buffer environment, the proteins were found to precipitate. In contrast, this was not the case with DENV 2 NGC RdRp. This observation also suggests that DENV1 and DENV 3 RdRp sensitivity to precipitation is likely contributed by the protein's

self-aggregation and amino acids composition. Further experiments like mutagenesis are needed to address this issue.

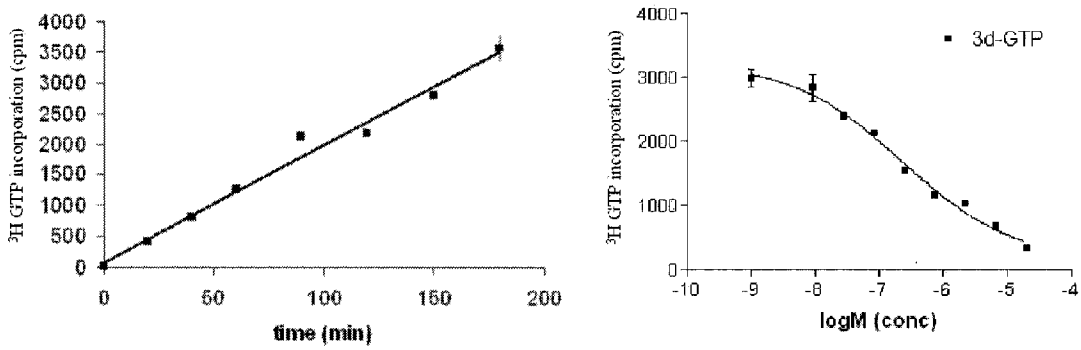


Fig 3.3.Kinetic of DENV 3 RdRp.

(A) Initial reading of incorporation within the linear range with respect to time by DENV 3 polymerase using poly C/primer oligoG template. The reaction mixture contains 1 μM enzyme, 0.25 μg poly C/0.0125 μg oligoG, 4 μM total GTP with 0.5 μCi of [^{3}H] GTP were described under Materials and Methods. (B) Kinetic analysis of 3'dGTP inhibition. The IC_{50} of 3' dGTP for DENV 3 RdRp was determined using at least two independent experiments. Assay conditions are described in the Materials and Methods section.

3.2.2 Fab antibody used for structural and functional studies

Protein crystallization is often inhibited by heterogeneity, insolubility, polydispersity or flexibility of a molecule in solution. To resolve these problems, antibody Fab or Fv fragments have been integrated as part of the strategy to crystallize such characteristic proteins (Kovari et al., 1995). The use of Fv fragments over Fab fragments has some advantages, as Fv fragment has no flexible elbow which may inhibit crystallization (Kovari et al., 1995). Nevertheless, Fab fragment has a reasonable solubility and binds specifically to an antigen.

Thus Fab antibody can effectively transform the aggregated material into soluble and monodisperse sample (Kovari et al., 1995), a prerequisite for crystallization. In addition, Fab antibody can serve to immobilize flexible regions of the target protein in a fixed conformation. The Fab-antigen interaction provides additional protein surfaces that facilitating nuclei and ordered crystal growth (Venturi and Hunte, 2003). This approach can be applied readily more than molecular engineering (point mutations or truncations) to obtain crystals with improved protein. HIV capsid protein p24 is that of an early example of insoluble protein which was crystallized with Fab (Prongay et al., 1990). This capsid protein p24 associate into large oligomers and this property interrupt attempts to crystallize the protein even with the presence of detergent ß-octyl-glucoside that improve the aggregation. The screenings of several different Fab antibodies that recognize p24 eventually led to encounter crystal conditions that suitable for diffraction. The use of Fab fragment to crystallize polymerase protein was first applied in HIV-RT, which was a ternary complex of RT-Fab-DNA (Jacobomolina et al., 1993). The Fab was suggested to act as a molecular clamp that immobilizes a region of the protein. Furthermore, the inclusion of Fab in HIV-RT-Fab-DNA complex resulted in rapid structure determination (Jacobomolina et al., 1993). To crystallize the FLNS5 protein, we adopted a similar strategies as that of HIV-RT or p24 (see section 3.3 for crystallization campaign on FL NS5), a human phage Fab library (1.2 $x10^{10}$) was biopanned to against DENV 2 FLNS5 (TSVO1 strain). After 3 rounds of biopanning and 2 rounds of ELISA screening, 5 Fab fragments that specifically recognize DENV 2 FLNS5 with binding affinity (K_d) in nM were selected and subcloned into vector for expression in *E.coli*. The K_d for respective Fab fragments were characterized by ELISA (see material and methods) and currently being validated by Biacore analysis (in collaboration with Dr. Susana Geifman, NTU). Further study on some of these Fab fragments that recognize the NLS region of RdRp (see below) are also in the progress to characterize the NS3-NS5 interaction

using Biacore approach. The NS3 was predicted to bind to 20 conserved amino acids in the region of bNLS (see introduction). The characterization of the Fab antibodies is discussed below (in collaboration with Dr. Celine Nkenfou, NITD, unpublished data). These Fab antibodies were highly expressed in TG1 *E.coli* (Fig 3.4, panel i), yielded about 5-10 mg/L culture and their features were summarized in Table 3.3.

Table 3.3

Antibody: Fab	**characteristic**	**K_d (nM)**
A8	-possible epitope: residues 359-415 -linear epitope	N.A
B5	-epitope uncharacterized -linear epitope	0.1
D12	-possible epitope: residues 359-415 -linear epitope	1.2
E6	-N.D in immunoblotting -conformational epitope	0.01
C3	-N.D in immunoblotting -conformational epitope	4.3

Table 3.3.Characteristic of Fab recognize DENV 2 NS5.

The table also shows their binding affinity which determined by ELISA assays (provided by Dr. Celine Nkenfou, NITD, unpublished data).

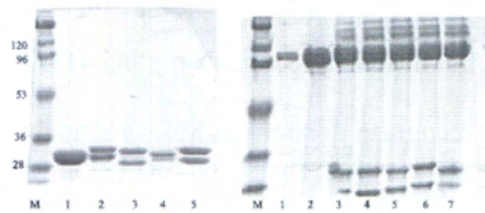


Fig 3. 4.SDS PAGE gel analysis.

(Left): profile for each mAb Fab that expressed in *E.coli* TG1. Lane 1-5 represents Fab D12, C3, E6, A8 and B5 respectively. (Right): DEN2 FLNS5-Fab complex formation in buffer without β-ME. Lane 1-2, DENV 2 FLNS5 in buffer without β-ME. Lane 3-7 represents FL NS5-Fab complex formation with aggregates (on top of the gel) for respective antibodies; follow the order on left hand side gel. These aggregates could be removed by gel filtration.

3.2.2.1 Epitope Mapping

Soluble protein expressions which led us to have proteins for Fab fragment characterization were discussed in earlier section. In this work, 12 proteins were chosen for 5 Fab fragments epitope mapping, includes the truncations or FLNS5 protein, which either from the DENV 2 serotype (TSVO1 or NGC strain), DENV 1 or 3 serotype and WNV NS5 (Fig 3.5 and Table 3.2).

However, these Fabs could recognize DENV 2 NS5 (TSVO1 strain) only, even though DENV 1-4 NS5 share more than 70% protein similarity. Three Fab antibodies of that -B5, D12 and A8 show strong reaction both in Elisa and immunoblotting, suggesting that these Fab fragments recognize a linear epitope (Fig 3.6). In contrast, C3 and E6 Fab only recognized DENV 2 FLNS5 in ELISA, suggesting that these Fabs recognize a conformation-sensitive epitope (Fig 3.7). For linear epitope mapping, immunoblotting and ELISA assay show that Fab A8 and D12 recognizes truncated proteins of that 341-900, 359-866 and FLNS5 respectively, but unable to recognize protein 415-900 and 48-359. The results suggest these Fab fragments recognize neither the MTase domain (1-296aa) nor residues after 415. This led us to propose that the epitope for Fab A8 and D12 lies within the residues 359-415 of the NLS regions. Sequence alignment (Fig 3.8) shows the most variable region in the NLS are starting at residue 366 to 392 (region A), with a few conserved residues.

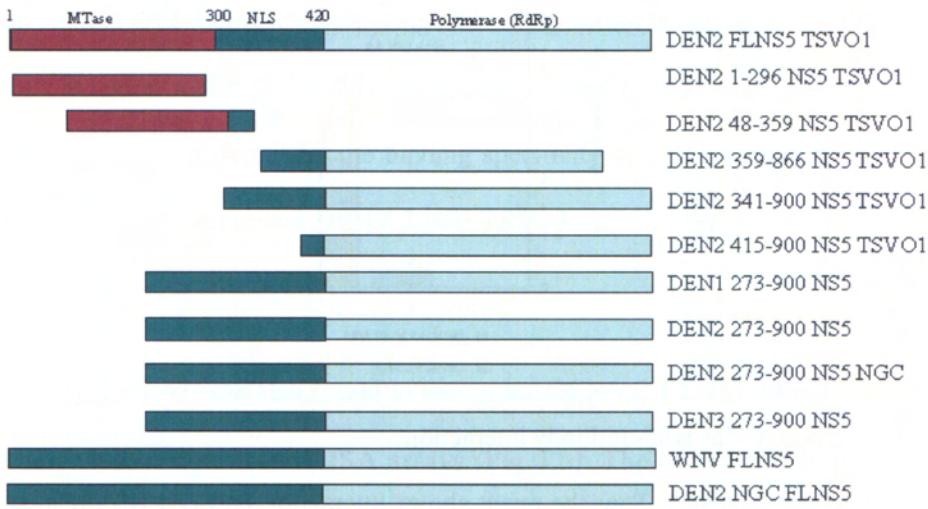
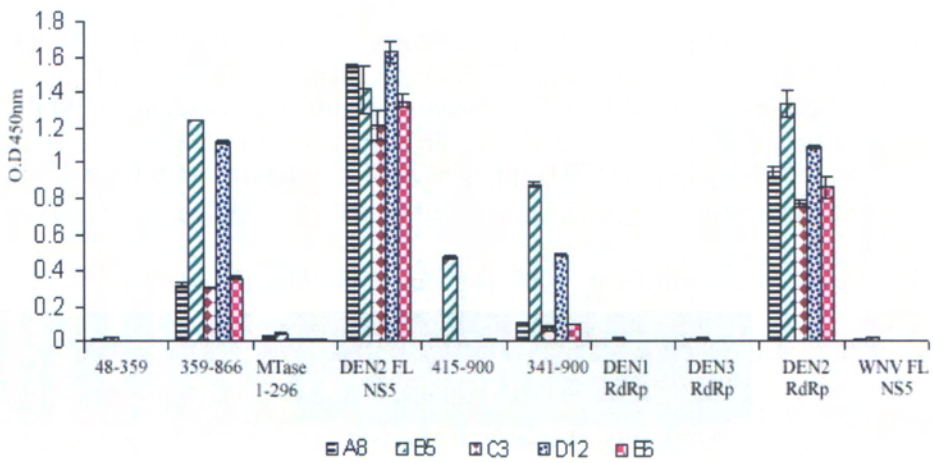


Fig 3.5.Schematic representation of the constructs of NS5.

A total of 12 NS5 related proteins were used for Fab fragments epitope mapping. For FLNS5, WNV, DENV 2 TSVO1 and DENV 2 NGC strain were used for mapping study. While the RdRp catalytic domains that have been chosen for mapping were DENV 1, DENV 2, DENV 2 NGC and DENV 3. The rest of the proteins were derivatives of TSVO1 strain. All the Fab fragments recognized DENV 2 TSVO1 but not any other DENV serotypes or WNV. Fab fragment of A8 and D12 were found to recognize residues 359-415 of NLS region. The MTase, NLS and polymerase (RdRp) regions are respectively represented by pink, dark green and light blue color.



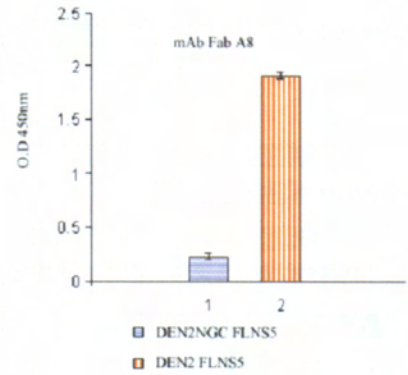


Fig 3.6.Elisa Assays for NS5-antibody interaction.

(Top)Elisa assays analysis. Top: the result shows interaction between NS5 related proteins (see Table 2 & Fig 3.4) and 5 Fab fragments of that A8, B5, C3, D12 and E6. See text for detail. (Bottom) the result shows the Fab A8 binding level between the DENV 2 NGC FLNS5 and DENV 2 FLNS5. See text for detail.

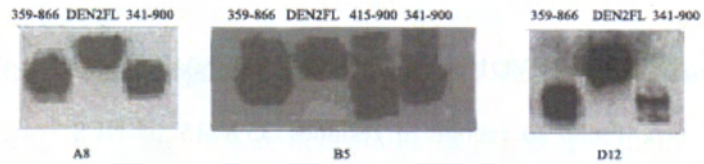


Fig 3.7.Western blot analysis.

Immunoblotting for Fab that recognizes DENV 2 NS5 related proteins (TSVO1 strain), respectively. No detection was found for protein DENV 1 RdRp, DENV 3 RdRp (273-900), WNV FLNS5 or domain III E protein (negative control, data not shown). Whilst Fab E6 and C3 reacted with DENV 2 NS5 (TSVO1) proteins in ELISA assays. See text for detail.

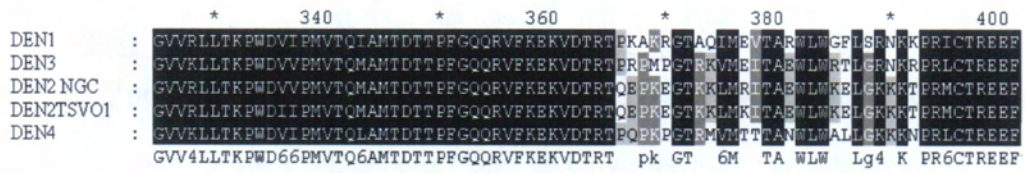


Fig 3.8.Sequence alignment analysis of NLS region from DENV 1-DENV 4.

The Fab A8 and D12 fragments were mapped binding within the residue 359-415 of NLS region. The variable region within the NS5 NLS region is starting from residue 376-392 (region A). The region A is likely to differentiate the binding specificity of Fab A8 and D12 to DENV1-4 and therefore specifically recognizes DENV 2 NS5 TSVO1.

Interestingly, Fab A8 was also found to recognize DENV 2 FLNS5 NGC 10 folds lesser than DENV 2 FLNS5 TSVO1 in ELISA assays (Fig 3.6). The others Fab fragments recognized DENV 2 FLNS5 NGC similar to that of TSVO1 strain (data not shown). The difference between the NGC and TSVO1 NS5 in residue 359-415 of NLS region are at residue 377 (Arg-->Lys) and 403 (Lys-->Arg) (Fig 3.8). Therefore, we mutated these residues on NGC NS5 plasmid to TSVO1 NS5 residues. However, Fab A8 binding on these mutants were as poor as the WT NGC NS5. As the interface between Fab and antigen is usually very large (>1000Å^2) (Evans et al., 2005), we thus suggested that residues reside out of the region A (see Fig 3.8) but folds into proximity of region A could also affect Fab A8 binding. Since our aim was to focus on Fab co-crystallization, residues that involves in respective Fab epitope was not subjected for further investigation.

In the case of conformational epitope, both Fab C3 and E6 that recognized the conformation epitope show that DENV 2 FLNS5 (TSVO1) has ~20% reaction stronger than of the DENV 2 RdRp (TSVO1, residues 273-900). These Fab fragments also weakly recognized proteins that contain partial NLS (protein 359-866 and 341-900) in ELISA assay. In combination of enzymatic activity, limited proteolysis, and characterization of C3/E6 epitope, this work suggested that RdRp in the FLNS5 context and RdRp alone adopt a different conformation.

How the role and function of RdRp is modulated by the presence of MTase and NLS region in the context of FL NS5 remains ambiguous.

3.2.2.2 Complex formation

Analytical size exclusion chromatography (Tricorn High Performance columns, 10/300 GL) has been used to study conditions optimal for FLNS5-Fab complex formation and co-crystallization screening. The Fab antibody has heavy chain and light chain which is linked with disulfide bonds. Therefore Fab may be sensitive to reducing agents such as DTT or β-ME. In this study, the FLNS5-Fab complex formation in solution without reducing agents resulted in aggregate formation, as shown by SDS PAGE gel and gel filtration (eluted at 7.5 ml, Fig 3.9, panel iii). However, the aggregate could be removed by size exclusion chromatography. Fab C3 was chosen for co-crystallization with FLNS5 in a buffer without β-ME. No crystal growth condition was identified at the moment. To improve the complex formation, we explored the use of β-ME. The result showed most Fab remain as in intact Fab in 2 mM β-ME (Fig 3.9, panel ii) and the Fab-FLNS5 complex was eluted as a monodisperse fraction (Fig 3.9, panel iv). This optimized buffer can be used for co-crystallization screening in the future. Similarly, the His tagged Fab fragments could be trapped in the His tag column. This column can then be used to co-purify with the untagged FLNS5, as a step to polish the complex before size exclusion chromatography purification and crystallization screening. In addition, biacore analysis on the Fab and FLNS5 interaction also can derives parameter, for example the K_{on} and K_{off} kinetic rate which could be taken into account when setting up crystallization trials. The rapid dissociation rate between Fab-FLNS5 can reduce the success rate for crystallization screening.

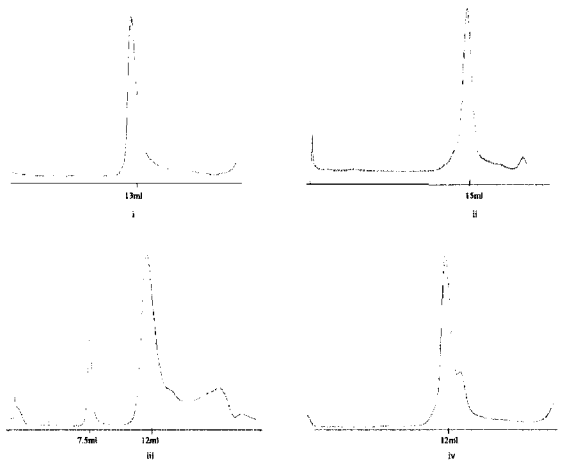


Fig 3.9.Representative gel filtration analysis for DENV 2 FL NS5 and Fab.

Analytical gel filtration (Tricorn™ High Performance Columns (10/300 GL) was used. Panel (I): profile for DENV 2 FL NS5, eluted at 13 ml. Panel (II): profile for Fab C3 treated with 2 mM β-ME remains intact and eluted at 15 ml. Panel (III): profile for DEN2 FLNS5-Fab C3 complex in buffer contains no β-ME. Profile shows two products, with complex eluted at 12 ml and aggregates at 7.5 ml Panel (IV): profile for the FLNS5-Fab complex in buffer with 2 mM β-ME could avoid the aggregates formation and the complex eluted at 12 ml.

3.2.2.3 Fab and enzymatic activity DENV NS5 RdRp

As shown in Fig 3.10, the polymerase activity of DENV 2 FLNS5 (TSVO1 strain) was not significantly affected by increasing the concentration of respective Fab fragments. This suggests that the binding of the Fab fragments on polymerase enzyme does not interfere with the polymerization catalysis by the enzyme. Therefore, the Fab could be potentially used to capture the active NS5 polymerase from the viral infected cells. The isolated complexes can be used for direct measurement of catalytic activity as secondary or tertiary drug screening assays or applied for structural studies. For example, anti influenza polymerase-antibody has been used to capture the enzyme. The complex was structurally determined by cryo-electron microscopy (Area et al., 2004). Similarly, the Fab should be useful in helping t to solve the

FLNS5 structure in solution using approaches such as small angle scattering, SAXS (is discussed in conclusion and outlook).

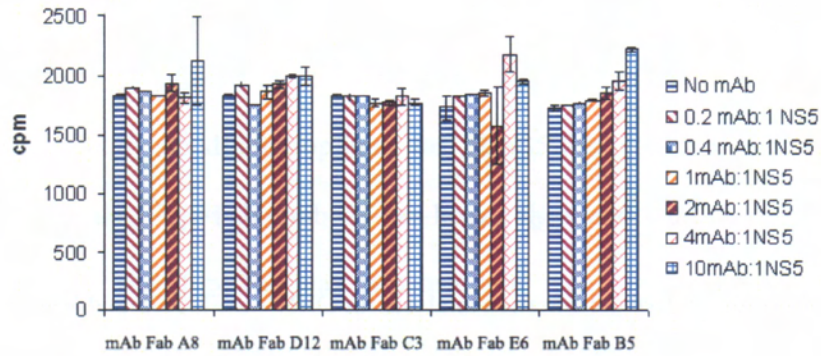


Fig 3.10.Enzymatic activity assays for DENV 2 FL NS5

The enzymatic assays were tested with the presence of respective Fab. The Fabs have no significant effect on NS5 polymerase catalytic activity. Increased polymerase activity was found for certain Fabs that have high Fab ratio NS5 (such as B5 and A8). Assay condition was tested as described in material and methods.

3.3 Crystallization assay

A total of 25 NS5 related proteins were constructed, expressed and screened, 13 of them were useful and selected for functional characterization, but only 6 of them were soluble, in good yield, stable and homogenous in solution, the prerequisite conditions for crystallization assays (Table 3.2). These 6 proteins were (1) DENV 2 FL NS5 expressed using baculovirus expression, (2) MTase DENV 2 1-296, (3) DENV 2 FL NS5 NGC, (4) WNV FL NS5, (5) DENV 1 RdRp (C-terminal His tagged) and (6) DENV 3 RdRp. The later 3 proteins were derived from homologous protein screening. For crystallization screenings, either FLNS5 proteins or the RdRp proteins were used:

(A) Several approaches were undertaken to crystallize the DENV 2 FLNS5 protein. Firstly, stabilizing agent osmolytes such as amino acids like glycine, proline and detergent ß-octyl-glucoside were included in crystallizing the protein after initial screenings failure. Osmolyte agent was also used as was shown in lysozyme such that osmolytes stabilize the protein structure against thermal denaturation, acting like osmotic-pressure to keep the protein stable (Arakawa and Timasheff, 1985). In addition, ß-octyl-glucoside was shown to decrease aggregation and help in membrane proteins crystallization (Modis et al., 2003). Secondly, crystallization assays with the WNV FLNS5 or WNV FLNS5 mutant (see appendix A2) were also set up, hoping the homologous proteins of DENV NS5 could yield crystals. Thirdly, taking a different approach, attempts to crystallize NS5-Fab complex are in progress. The Fab antibodies specifically recognize DENV 2 FLNS5 (see section 3.2). Fourthly, NS3 helicase has been reported to interact with NS5. Co-crystallization of WNV FLNS5-FLNS3 is also in progress to increase the likelihood of obtaining crystals. Initial ELISA assays in this study shows that NS5 could interact with NS3 (Fig 3.11), and is consistent with biacore NS5-NS3 interaction studies (data not shown, in collaboration with Dr. Susana Geifman, NTU). Nevertheless, in summary none of these approaches resulted in crystals. These results indicate the challenges of crystallizing the FLNS5 proteins which probably may be due to the flexibility and mobility between the MTase and RdRp domain.

(B) Homologous protein screening of catalytic domain RdRp (273-900) identified DEN3 RdRp as the best soluble protein. For the domain protein crystallization screening (DEN1 & 3 RdRp, 273-900), one condition that produced DENV 3 RdRp crystals was encountered. The detail of crystal optimizations is discussed in chapter 4. On the other hand, DENV 2 MTase 1-296 protein (Fig 3.2d, lane 7; protein 19 Table 3.1) was also expressed and purified. The

aim was to use the reproduced the crystals to serve for crystal-inhibitor screening. However, work of Egloff et al., 2002 was not reproducible in our lab.

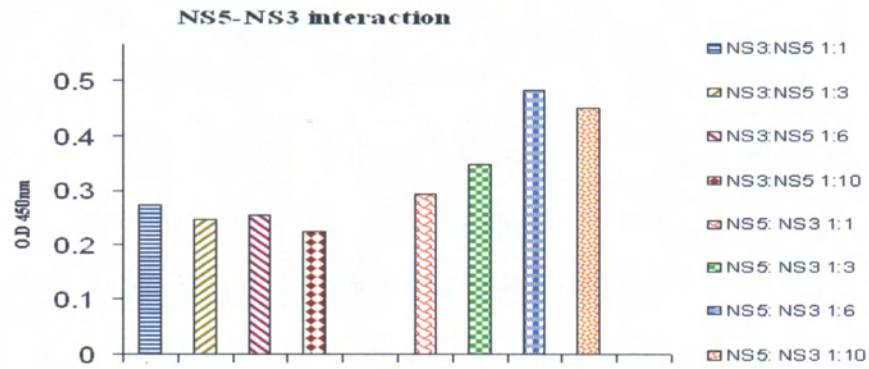


Fig 3.11.ELISA assays analysis of NS5-NS3 (WNV) interaction.

Coated NS3 (bar 1-4) could interact with NS5 and reversely coated NS5 (bar 5-8) also interact with NS3. In both cases, BSA was used as control with respective ratio used in NS3-NS5 (or reverse) interaction. Increase NS5 to coated NS3 on ELISA plate to 10 NS5:1 NS3 has no effect on the interaction intensity (bar 1-4). In contrast, increase NS3 to coated NS5 on ELISA plate could slightly enhance the interaction (bar 5-8). For detection, NS5 was recognized by anti-NS5 polyclonal rabbit IgG and NS3 was recognized by anti-NS3 polyclonal mouse IgG, with ratio 1:5000.

3.4. Conclusion

In the study reported in this chapter, we have explored strategies that can lead to soluble NS5 protein expression. This work has designed 26 NS5 related protein constructs. After expression, purification and characterization, a total of 6 NS5 related proteins were found to satisfy the prerequisite condition of crystallization screening and also deemed useful for enzymatic assay studies. The six proteins are DENV 1 RdRp, DENV 3 RdRp, WNV FLNS5, DENV FLNS5 (TSVO1 strain), DENV 2 FLNS5 (NGC strain) and MTase DENV 2. Initially,

this work started with FL DENV NS5 (TSVO1 strain) proteins which were expressed in baculovirus system. The proteins were used for functional studies (eg.enzymatic assay or HTS campaign) and crystallization screening. However, extensive crystallization efforts have not yielded any crystals (see below). Thus, protein refolding, limited proteolysis and homologous protein screening were used to design truncated constructs that could be expressed in *E.coli* in order to increase the likelihood of obtaining crystals. With the trials of crystallization assays on the 6 proteins mentioned above, one crystal growth condition was encountered for DENV 3 RdRp. The initial crystals were small and diffracted poorly. See chapter 4 for strategies leading to high resolution and good quality crystal diffraction. This led us to have structural insights on the DENV RdRp structure (see chapter 5).

In our attempts to crystallize FLNS5, several strategies were employed. A homologous protein of DENV, the WNV FLNS5 was used for crystallization trials. WNV FLNS5 mutants that were predicted based on surface entropy reduction (see appendix A2) were also designed and used for crystallization trials. These attempts have not yielded any crystals, NS5-complex co-crystallization was employed. The WNV FLNS3 that was predicted to interact with NS5 during the replication was used for co-crystallization. Unfortunely, these attempts did not materialize to crystallize the protein. Taking a different approach, a Fab monoclonal antibody (mAb) library was biopanned against DENV 2 FLNS5 (TSVO1 strain). Bound phage clones of DEN2 FLNS5 were enriched and selected from biopanning and ELISA screening. Subsequently the best 5 Fab fragments were subcloned into expression vector. These the 5 Fab fragments that are specifically against DENV 2 FLNS5 were selected for crystallization attempts. Nevertheless, in summary all of our attempts have been so far unsuccessful.

For functional characterization, the enzymatic assay demonstrated DENV 3 RdRp that used for structural determination is an active protein and 3'dGTP (chain terminator) can serve as useful reference inhibitor. The sensitivity and speed of the assay make it suitable for high-throughput screening of inhibitors. In addition, a serendipitous result also reveals that DENV 2 RdRp NGC is ~4 folds less active than DENV 2 FLNS5 (TSVO1 strain), compared to that of DEN3 RdRp which is ~20 folds less active than the FL protein. The results indicate that the catalytic domain RdRp protein adopts a more open conformation and is more flexible than that of FLNS5. This suggestion is supported by the epitope mapping and limited proteolysis analysis. In addition, the variation of amino acids sequence in DENV 1-3 RdRp also could affect the RdRp and indeed DENV 2 RdRp (NGC strain) activity is more active than that of DENV 3 RdRp. In the last part of this work, 5 Fab fragments that recognize DENV 2 NS5 were characterized. Three Fab fragments were shown to recognize linear epitope while two Fab fragments were found to recognize conformational epitope. None of these mAbs affect the enzymatic activity of polymerase. Nevertheless, these Fab fragments should serve as good tools to study the solution of FLNS5 structure using SAXS in the future.

CHAPTER 4

A MULTI-STEP STRATEGY TO OBTAIN CRYSTALS OF THE DENGUE VIRUS RNA-DEPENDENT RNA POLYMERASE THAT DIFFRACT TO HIGH RESOLUTION AND SOAKING STUDIES

INTRODUCTION

A well ordered and high quality crystal is pre- requisite for data collection and determining a protein structure. A beautiful looking crystal does not imply good diffraction quality (Owen and Garman, 2005). Loosely packed molecules and a large solvent volume are the common factors that result in poor crystal diffraction and low resolution. In the case of well grown crystals that do not diffract, generally the approaches are to search for new crystallization conditions to identify new crystal form, crystallizing a truncated form of the protein or mutating surface amino acids to enhance protein crystallization (D'Arcy et al., 1999). Crystal growth remains an empirical and tedious process. Sometimes minor modification such as surface residues mutation on the same protein may not produce crystal under similar conditions as before. Thus, several methods should be considered to improve poorly diffracting crystals include post crystallization soaking, cross-linking, crystal annealing, and controlled dehydration (Heras and Martin, 2005). These methods have been reported to improve diffraction of protein crystals. In this study, we have identified one crystal growth condition for DENV 3 RdRp, the best protein construct with respect to the protein solubility, monodispersity, and stability (see chapter 3). However, the crystal diffracted poorly.

Therefore we focused on the DENV 3 RdRp crystal quality and its diffraction property improvement, in order to understand its function and role from structural basis.

Objective of the study

Viral polymerases represent attractive drug targets for the development of specific drugs, as host cells are devoid of this enzymatic activity. To accelerate the search for novel antiviral compounds that are active against DENV, we explored the multi-step strategy used which led to a good-resolution structure (1.85 Å) of the DENV 3 RdRp catalytic domain that suitable for further structure-based drug design. The strategy, outlined in Fig 4.1, involved (i) expression screening of naturally occurring serotype variants of the RdRp domain, (ii) a crystal dehydration/cryoprotection protocol and (iii) optimization of crystal-growth conditions. In addition, soaking conditions and substrates for complex studies is also briefly discussed.

RESULT AND DISCUSSION

4.1 Protein expression and purification

The expression and purification strategy for DENV 3 RdRp was discussed in chapter 2 and chapter 3 respectively (refer appendix A1). This protein was expressed as a soluble protein from the homologous protein screening strategy. Briefly, the proteins were purified using metal affinity, cation exchange and gel filtration chromatography sequentially. DENV 3 RdRp proteins retaining their N-terminal His_6 tag hereafter named H_6 DENV 3 RdRp.

In the case of H_6 DENV 3 RdRp, two distinct fractions (labelled F1 and F2) were eluted from cation-exchange chromatography (Fig 4.2a) and were separately purified by gel-filtration chromatography (see below). Alternatively, after elution from the HisTrap HP column the H_6

DEN3 RdRp protein was treated with thrombin overnight at 277 K in order to remove its N-terminal hexahistidine (His_6) tag using 1 U thrombin per 100 µg of protein. After treatment with thrombin, the DENV 3 RdRp protein retains residues GSHMLDN at its N-terminus which are derived from the vector that was used for cloning. Finally, both the tagged and untagged proteins were concentrated by ultrafiltration with a molecular-weight cutoff of 30 kDa (Millipore, Volketswil, Switzerland), a final gel-filtration chromatography step (HiPrep 16/26, Superdex 200) was carried out in buffer C [20 mM Tris-HCl pH 6.8, 0.25 M NaCl, 1 mM EDTA, 2 mM β-mercaptoethanol and 0.1%(w/v) CHAPS] and both the tagged and untagged proteins were concentrated to ~11 mg ml^{-1} and used in crystallization assays.

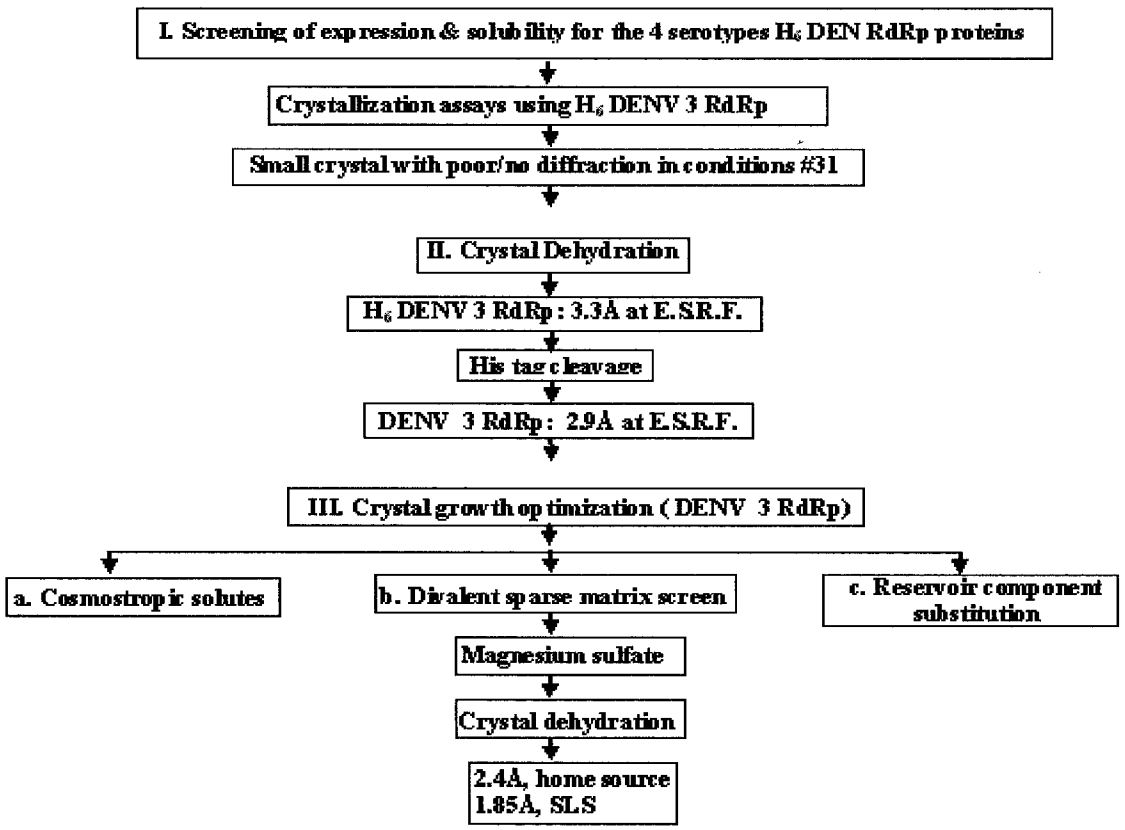
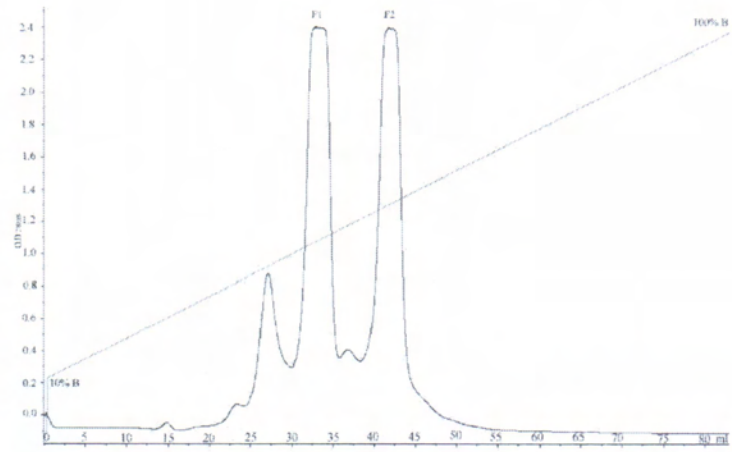


Fig 4.1.Flow-chart for strategy to obtain good resolution structure.

The approach consisted of the following steps (I) The 4 serotypes of the target protein (containing a hexahistidine tag) were screened for expression levels and solubility, with the RdRp from serotype 3 giving the best results (see chapter 3 for detail) (II) A major improvement of the diffraction quality for small crystals (obtained using conditions #31) was obtained by air dehydration. Removal of the N-terminal His_6 tag resulted in better diffraction. (III) Crystals of DENV 3 RdRp (with no His_6 tag) were further optimized using magnesium sulfate or magnesium acetate as additives.

A



B

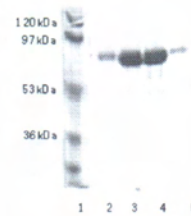


Fig 4.2.Purification profile and SDS PAGE analysis.

(A) Elution profile from the cation exchange purification step for the H_6 DENV 3 RdRp protein. The linear gradient ranges from 0.15 M NaCl (corresponding to the inclusion of 10% of Buffer B)

to 1.5 M NaCl (100% B). Two major fractions F1 and F2 were then subjected to gel filtration and initial crystallization trials (see text). (B) SDS PAGE analysis. Lane 1, molecular mass markers. Lane 2 & 5, protein DENV 3 RdRp from the minor fraction that elutes first. Lane 3 & 4, fractions F1 and F2 respectively both containing DENV 3 RdRp.

4.2 Crystallization

The H_6 DENV 3 RdRp protein was subjected to initial crystallization assays, using the F1 and F2 fractions eluted from the cation exchange chromatography step separately (Fig 4.2). Examination of the initial crystallization assays led to the following observations: trials performed at 293 K gave protein precipitates in almost all conditions investigated. Therefore, subsequent crystallization assays were performed at 277 K. Using the F1 fraction of the H_6 DENV 3 RdRp and conditions #31 of the Index Screen from Hampton Research (0.1M Tris-HCl ph 8.5, 0.8M potassium/sodium tartrate, 0.5%(w/v) PEG MME 5000), small crystals (Fig 4.3, panel I) with approximate dimensions up to of 0.05 × 0.02 × 0.02 mm^3 were obtained that diffracted X-rays very weakly. Interestingly, no crystal could be obtained using the same crystallization conditions with the F2 fraction or pools of fractions F1 and F2. Since the protein DEN3 RdRp which is devoid of the His_6 tag elutes as a single peak after ion-exchange chromatography, fractions F1 and F2 at pH 6.0 might be caused by the presence of the N-terminal His_6 resulting in the formation of two conformational states and two major elution peaks (Fig 4.2).

4.3 Post crystallization treatment

It was suspected that poor diffraction was provoked by extensive dynamic properties of the enzyme, introducing disorder in the crystal lattice, as observed previously for crystals of the HIV-1 reverse transcriptase (Esnouf et al., 1998). In some cases, poor diffracted crystals

could greatly improve the diffraction resolution and quality after treatment with annealing. Unfortunately it was not the case for DENV 3 RdRp. Thus, we established a careful air dehydration procedure in the hope that the provoked reduction of solvent content could produce a more closely packed and better ordered lattice. By contrast with the HIV-1 RT, we were not able to detect changes triggered by crystal dehydration in the unit cell parameters for H_6 DENV 3 RdRp crystals, since initial diffraction was too poor to determine unit-cell dimensions. Various PEG solutions with average molecular masses ranging from 200 to 3350 were tested as potential dehydration/cryoprotection agents, as listed in Table 4.1. From this preliminary screen, PEG 200, PEG 400 and PEG 3350 were the most promising because they did not provoke any visible damage to the crystals.

A summary of the conditions tested and the resulting improvements in diffraction obtained is given in Table 4.2. The optimal dehydration/cryoprotection protocol involved the serial transfer of crystals at 277 K in volumes of 5 µl of the precipitating solution supplemented with increasing concentrations of PEG 3350, ranging from 15%, 25% to 30% (w/v), with incubation times of 15 min, 15 min and 30 min respectively. In order to achieve a final concentration of 30% (w/v) PEG 3350 for the final dehydration step, the concentration of the Na/K tartrate was lowered from 0.8 M to 0.3 M to avoid solubility problems.

Crystals for H_6 DENV 3 RdRp subjected to this procedure diffracted to ~4Å on a radiation home source and to 3.3 Å resolution at the ESRF, belong to space group $C222_1$ and have average unit-cell parameters of a= 160.28 Å, b= 178.77 Å, and c= 58.05 Å, with only slight variations depending on the dehydration agent. Crystals for the DENV 3 RdRp protein (with the His_6 tag removed through thrombin cleavage) could also be obtained using the same condition #31. These crystals, which were subjected to the same dehydration/cryoprotection

procedure outlined above, diffracted better (to 2.9 Å at the ESRF ID14-4 beam-line) than those originally obtained with the H_6 DENV 3 RdRp protein (to 3.3 Å at ESRF) (Table 1 & Fig.1). Thus further crystal optimizations made use of the DENV 3 RdRp protein devoid of the histidine tag.

Table 4.1

Dehydration agents	Concentrations (w/v)[a]	Resolution, Å	Radiation source
No dehydration	-	~20	home source
PEG 200	10%, 15%, 35%	~ 7	home source
PEG 300	10%, 15%, 25%	~10	home source
PEG 400	10%, 15%, 25%	~5	home source
PEG 600	10%, 15%, 25%	~10	home source
K/Na tartrate[b]	0.84 M, 0.9 M, 1.0 M	~15	home source
MPD	(dissolves crystals)	-	home source
Glycerol	10%, 15%, 30%	~ 8	home source
PEG 3350[c]	15%, 25%, 30%	4.5 [d] 2.9 [d] 2.4[e] 1.85[e]	home source synchrotron home source Synchrotron

Table 4.1.Summary of dehydration/cryoprotection assays on DENV 3 RdRp crystals.

[a] Air dehydration involved the stepwise transfer of crystals into a 5 µl drop for 15 min, 15 min and 30 min respectively in each solution. Each drop contains the original precipitating solution #31 (see Methods) with the addition of the dehydration agents listed above. Longer durations of dehydration yielded no improvement in crystal diffraction, (eg stepwise transfers for a total time of 24 h). Very brief soaking times (~10 s) did not improve diffraction. [b] In this case, the cryoprotectant is a solution containing 30% (w/v) glycerol after crystals had been dehydrated. [c]

Optimal dehydration/cryoprotection used PEG 3350 for DENV 3 RdRp, leading to high resolution diffraction.

4.3 Further Crystal optimization

In order to improve diffraction properties further, several strategies were iteratively experimented, including the use of "cosmotropic" solutes, the addition of co-factors and substitution of individual components within the reservoir solution (Fig 4.3 and Table 4.3). All crystal optimization assays were conducted using condition #31 as a starting basis, with the agents listed in Table 4.2. The effects of each factor are summarized below and typical crystals obtained using the various procedures are displayed in Fig 4.3.

4.3.1 Cosmotropic solutes

Cosmotropic solutes, such as glycerol or glucose are agents that preserve protein structure against thermal denaturation or reduce conformational heterogeneity. This approach has been successfully applied to generate crystals suitable for X-ray diffraction studies in several cases, including the T7 RNA polymerase (Sousa et al., 1994) or a human fibronectin fragment (Pechik et al., 1993), where addition of glycerol proved crucial to improve crystal quality. Several cosmotropic solutes were tested and the results are summarized in Table 4.2 and Fig 4.3. Overall, in our case, these additives either prevented crystal growth (eg: iso-propanol) or accelerated the rate of nucleation and crystal growth, producing smaller crystals as shown in panel IV of Fig 4.3.

4.3.2 Role of divalent ions

The DENV RdRp requires divalent ions (manganese or magnesium) for its enzymatic activity. Several salts of divalent ions (as listed in Table 4.2) were used as additive to improve crystal

quality. Surprisingly, crystallization condition # 31 supplemented with concentrations of 2 mM of manganese chloride, magnesium chloride or 5 mM calcium chloride proved detrimental to crystal formation (Table 4.2 and Fig 4.3 panel IV). By contrast, addition of either magnesium acetate or magnesium sulfate at various concentrations resulted in a strikingly improved crystal quality (Table 4.2, Fig 4.3 and 4.4). Explanations for this effect might be found by reference with previous studies carried out on lysozyme solubility and crystal growth using various ions (Rieskautt and Ducruix, 1989). The best conditions found made use of 5 to 40 mM magnesium sulfate as an additive to condition #31, leading to crystals diffracting to resolutions of 1.85 Å at SLS (Fig 4.4) and to 2.4 Å (Table 4.2) on a rotating anode, following the procedure outlined in Fig 4.1 and described above. These crystals were used for structure determination and refinement.

4.3.3 Replacement of individual component within the reservoir

Substitution of PEG 5000 monomethyl ether with various other PEG (while retaining the other components of the crystallization solution in the reservoir (Table 4.2) produced two distinct results. Solutions containing PEG 200 or PEG 400 (with concentrations of 2.5 to 5 % (w/v) produced large crystals comparable to those displayed in panel II of Fig 4.3, but these crystals grow more slowly, in approximately 10 weeks. With reservoirs containing PEG 1500 or PEG 2000 (1-2.5 % w/v), crystal formation occurred within 5-6 days (panel III), with a size/morphology very similar to crystals obtained using the original conditions. These results show that PEG of lower molecular mass (PEG 200 or PEG 400) slow nucleation as compared to PEG 1500 or PEG 2000, which accelerate crystal growth.

Table 4.2

Crystal optimization	Observation
Condition #31 (original precipitating solution)	outcome I
I. Cosmotropic screen (Jeruzalmi & Steitz, 1997): ethylene glycol, sucrose, glycerol, iso-propanol, &glucose (w/v)	
ethylene glycol	
5%	precipitation
10%	outcome I
15%	deter crystal growth
sucrose	
5%, 10%, 15%	outcome IV
glycerol	
5%, 10%	outcome IV
15%	deter crystal growth
iso-propanol	
5%, 10%	deter crystal growth
15%	precipitation
glucose	
5%, 10%	outcome IV
15%	outcome I
II. Sparse matrix co-factor screen (divalent ions) (Boddupalli et al., 1992, Doudna et al., 1993)	
5mM $CaCl_2$ 2mM$MgCl_2$	outcome IV
0.2M $MgCl_2$, 2mM $MgCl_2$ or 2mM $MnCl_2$	deter crystal growth
5mM $MgCl_2$ 2mM $MnCl_2$ or 2mM MnCl2 2mM $MgCl_2$	
5mM $MgSO_4$, 40mM $MgSO_4$, 0.2M $MgSO_4$ or 0.2M magnesium Acetate	outcome II (optimal conditions)
III. PEG substitution in mother reservoir (crystallization drop using condition #31)	
PEG 1.5K and 2K (1- 2.5% w/v)	outcome III
PEG 200 and 400 (2.5- 5% w/v)	outcome II (Crystal formation >10 weeks)

Table 4.2.Summary of further crystal optimization assays for DENV 3 RdRp crystals. Please see text for a discussion on crystal growth optimization. The crystals obtained ("observed outcomes") are shown in Fig. 3. [a] Crystal optimization was conducted with condition #31 as a starting basis, supplemented with reagents listed in Table 3.

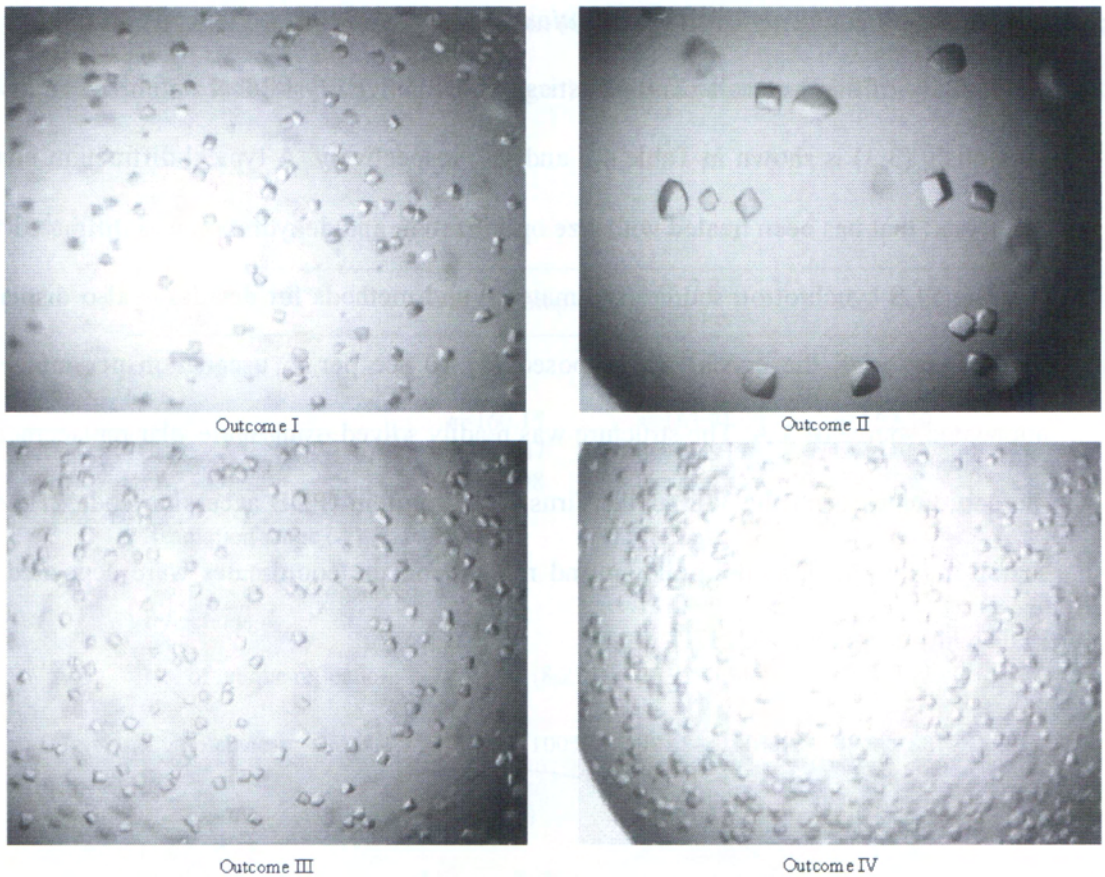


Fig 4.3.Typical crystals obtained during crystal optimization assays.

The same magnification of × 20 is used in all panels. Outcome I: small crystals from the original precipitating solution with linear dimensions < 0.05 mm appearing in ~3-4 weeks. Outcome II: crystal obtained with the addition of divalent ions. Significantly larger crystals appear in 3-4 weeks (see text). Outcome III: substitution with PEG 1500 and PEG 2000 in the reservoir leads to a more rapid growth of crystals without increasing their final size, which is attained in ~5-6 days. Substitution with PEG 200 and PEG 400 produced crystals like those shown in panel II. Outcome IV, increased rate of nucleation of crystals (approximately 1 week), leading to slightly smaller crystals (outcome I).

4.3.4 Data collection and structure solution

A summary of the data collection statistics for the native crystal that optimized as discussed previously (3.3) is shown in Table 4.3 and 4.4, respectively. A typical diffraction image for the crystal that has been treated with size optimization and dehydration was diffracted to 1.85 Å using SLS synchrotron source (see material and methods for details) is also displayed in Fig 4.4. In brief, the crystal was exposed for 10 sec per 1° osccilation per image using attenuated source of 1 Å. The structure was readily solved using molecular replacement with the coordinates from the West Nile Virus RdRp protein (PDB accession code 2HFZ) as a search model. After model building and refinement, the coordinates were deposited in the PDB with accession code 2J7U.

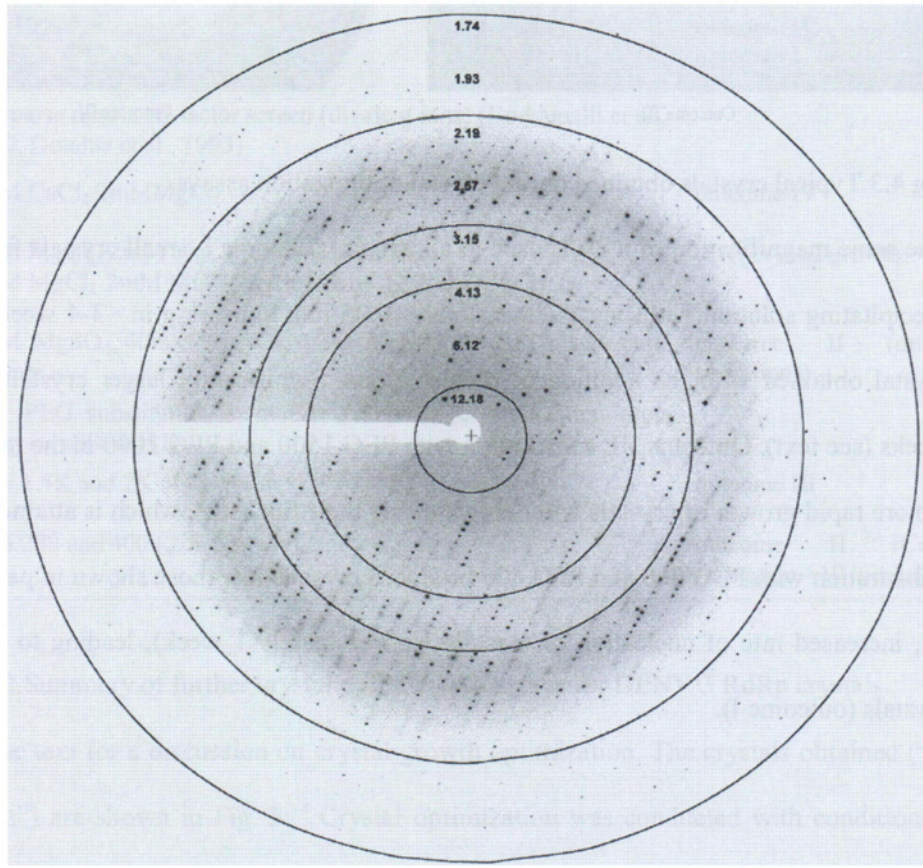


Fig 4.4.Typical Diffraction image.

Crystals of the DENV 3 RdRp catalytic domain (Outcome II) collected at a Synchrotron radiation source (Beamline X10SA, PXII, S.L.S. Villingen). The diffraction extends to 1.85 Å resolutions.

Table 4.3

	Mg^{2+} Complex	3'dGTP Complex	Mn^{2+}Complex
Wavelength (Å)	1.0	1.0	Home source
Cell parameters (Å), $C222_1$	a=160.15 b=180.47 C=57. 98	a=163.71 b=180.91 c=58.04	a=162.39 b=181.28 c=57.84
Resolution range (Å)	48-1.85	48-2.6	20-2.6
No. of observed reflections	355,151	131,992	93,113
No. of unique reflections[a]	72,065 (8,212)	26,929 (3067)	25,854 (3792)
Completeness	99.9% (100%)	99.6% (100%)	99.9%(98.7%)
Multiplicity	4.9 (4.9)	4.9 (4.9)	6.1 (5.2)
R_{merge}[b]	0.009 (0.540)	0.010 (0.627)	0.062 (0.294)
$I/\sigma(I)$	14.62 (2.98)	13.98 (2.54)	22.5 (5.60)

Table 4.3.Data Collection and phasing statistics.

[a]The numbers in parentheses refers to the last (highest) resolution shell. [b]$R_{merge} = \sum_h \sum_i |I_{hi} - \langle I_h \rangle| / \sum_{h,i} I_{hi}$, where I_{hi} is the ith observation of the reflection h, while $\langle I_h \rangle$ is its mean intensity.

Table 4.4

	Mg^{2+} Complex	3'dGTP Complex	Mn^{2+}Complex
Resolution range (Å)	20.0-1.85	20.0-2.6	37.0-2.60
No of reflections: completeness (%)	99.9	99.7	99.85
Used for refinement	68,360	25,549	25,472
Used for Rfree calculation	3,646	1,340	1,.348
No of non hydrogen atoms			
Protein	4,642(36.2)	4,621(62.4)	4,829(68.9)
missing residues	61	63	61
$Zn^{2+}/Mg^{2+}/Mn^{2+}$	2/1/0(28.4, 47.0, 0)	2/0/0(54.3,0,0)	2/0/2(37.1,0,56)
Water molecules			
Rfactor[§] (%)	19.6	21.3	22.7
Rfree# (%)	23.1	28.8	26.8
Rms deviations from ideality			
Bond lengths (Å)	0.007	0.010	0.006
Bond angles (°)	0.909	1.158	0.841
Ramanchandran plot data			
Residues in most favoured regions (%)	92.2	88.6	91.7
Residues in additional allowed regions (%)	7.0	10.2	7.8
Residues in generously allowed regions (%)	0.2	0.8	0.2
Residues in disallowed regions (%)	0.6	0.4	0.4
Overall G factor*	0.19	0.09	0.23

Table 4.4.Refinement statistics.

[§] Rfactor $=\sum||F_{obs}| - |F_{calc}|| / \sum |F_{obs}|$. [#] Rfree was calculated with 5% of reflections excluded from the whole refinement procedure.* G factor is the overall measure of structure quality from PROCHECK (Lee and Lobigs, 2000).

4.4 Soaking studies- conditions and substrates

In drug discovery, the structural information of a protein-inhibitors complex is used iteratively to design new molecules. With this aim, X-ray crystallography crystal-ligand screening was initiated in collaboration with NITD. A total of 22 inhibitors were selected (see appendix A3, for internal use only). Additionally, 2 different ligands (RNA template and Mn^{2+}) were also selected for soaking study in order to obtain the DENV RdRp-ligand complex which provide us clue to its function and role. The optimized soaking condition for respective ligands is shown in Table 4.5 and 4.6 and is discussed in section 4.4.1-4.4.3.

To obtain a protein-inhibitor complex, either through co crystallization or soaking is too often not successful, even when the crystallization conditions are well established for a protein of interest (Danley, 2006). Hit rate of 0.6-2% was reported in some recent fragment based screening studies (Murray et al., 2007; Bosch et al., 2006). Our attempts to identify inhibitor that binds DENV 3 RdRp crystal so far have been unsuccessful. However, the preliminary fragment based screening condition by DENV 3 RdRp X-ray crystallography studies could serve as a basic guide for the future investigation. There are several parameters affecting the hit rate, including protein-ligand binding equilibrium, kinetics of binding, ligand solubility, crystal lattice limitation, chemical or soaking environment. These factors should be taken into account in designing experiments (Hassell et al., 2007).

Table 4.5

No	pH	Salt	PEG(w/v)	Additive	Remarks	Diffraction(Å)
1.	90mM Tris HCl pH 8.5	80mM K/Na tartrate	20% PEG 3350	0.1M $MnCl_2$	Solution not soluble	N.A
2.	90mM Tris HCl pH 8.5	0.2M $MgCl_2$	25% PEG 3350		Crystal stable	2.5-2.6 Å
3.	90mM Tris HCl pH 8.5	0.3M $MgCl_2$	25% PEG 3350		Crystal stable	2.5-2.6 Å
4.	90mM Tris HCl pH 8.5	50mM $MnCl_2$	30% PEG 3350		Crystal stable, drop precipitated o/n	2.5-2.6 Å
5.	90mM Tris HCl pH 8.5	0.2M $CoCl_2$	20% PEG 3350		Crystal stable	Deters crystal diffraction
6.	90mM Tris HCl pH 8.5	0.2M $NiCl_2$	25% PEG 3350		Crystal stable	Deters crystal diffraction

Table 4.5.Summary of crystal soaking buffer optimization for DENV 3 RdRp crystals.

Table 4.6

Ligand	Soaking buffer	Comment
Inhibitors	90mM Tris HCl pH 8.5, 35% PEG 3350, 10mM $MgCl_2$	Crystals diffractable to 2.8-3.5Å at home source
RNA template- 5, 6, 7 mers dU	0.2M Tris HCl pH 8.5, 50mM $MgCl_2$, 30% w/v PEG 3350	No extra electron density map observed for 6 mers.
Divalent metal Mn^{2+}ion	90mM Tris HCl pH 8.5, 10mM $MnCl_2$, 25% PEG 3350	

Table 4.6.Summary of ligands used for soaking and their respective soaking optimization.

4.4.1 Stabilization of crystal

Protein crystals were grown in condition #31 (0.1 M Tris-HCl pH 8.5, 0.8 M potassium/sodium tartrate, 0.5% w/v PEG MME 5000). The mother liquor that contains 0.8 M potassium/sodium tartrate has two disadvantages. First, the addition of 35% w/v PEG 3350 could saturate the solution used for cryo-dehydration. Second, tartrate is a metal chelator similar to EDTA. Thus, this study has tried to search for artificial mother liquor to replace the tartrate and this is also equally important for the metal ions soaking study. A series of

artificial mother liquors were then tested. The best artificial mother liquor was that of condition #2 (90 mM Tris HCl pH 8.5, 0.2 M $MgCl_2$ and 25% PEG 3350, Table 4.5), with respect to its diffraction quality and stability. This solution was used to obtain RdRp-Mg^{2+} complex which diffracted to 1.85 Å (see chapter 5). In addition, this condition #2 was the basis parameter for inhibitor and RNA template soaking study (see below). Other conditions were detrimental to crystal diffraction or mother liquor drop precipitated after overnight (condition #4, Table 4.5). To stabilize crystals, they could be transferred stepwise over to the new artificial mother liquors by increased concentration of the precipitants or through introduction of a cryoprotectant to original crystal drop so that crystals are not damaged. In some cases, the crystal is cross-linked with glutaraldehyde to stabilize the crystal prior to carry out soaking experiments, as exemplified in the case cyclin A-cdk2 crystals (Hassell et al., 2007; Lusty, 1999).

4.4.2 Lattice compatibility and soaking time

Crystal packing in the lattice is one the parameters that has an effect on the interaction between protein and ligand. The crystal lattice must be compatible with the protein conformational changes after the ligand is bound (Hassell et al., 2007). The protein-ligand complex conformational changes that differs from the apo-enzyme, will lead to inability to grow crystal in the apo-enzyme condition (co crystallization) or causes damage to the crystal (soaking) during soaking. Fragment based screening with HCV polymerase has shown that non-nucleoside inhibitors do not induce huge conformational change that can disrupt the crystal lattice (Wang et al., 2003; Biswal et al., 2006). In addition, ligand-protein ratio is also an important parameter in co crystallization. Excess of ligand could deter crystal growth. Furthermore, optimal soaking incubation time can vary the ligand occupancy. In some cases, incubation time could range from several days up to 2-3 months to obtain the ligand into the

protein crystals of good occupancy. Stepwise crystal transfers from initial ligand drop to a newly ligand added drop for a period of time has been shown successful in some cases (Delarue et al., 2002). In the case of DENV 3 RdRp, ligand soaking studies were soaked for 16-48 h. Some exceptional ligands were soaked with crystals for up to 6 days, depending on the stability of the ligands and crystals (see appendix A3, for internal use only).

4.4.3 Inhibitor soaking

Most of the compounds (inhibitor) produced via medicinal chemistry have low aqueous solubility and need to be dissolved in organic solution such as DMSO (Lipinski, 2000). Extremely low aqueous solubility compound may encounter difficulty in soaking or co crystallization. In addition, the use of DMSO in protein-ligand complex has to be minimal to avoid damage on crystal or protein (Fig 4.5). Initially, some inhibitors tested in this study were insoluble in the condition # 2 (Table 4.5). In order to minimize the final concentration of DMSO while increase the inhibitor stock concentration which used in soaking, the Peg 3350 concentration was increased from 25% to 35%. Increasing PEG or alcohol in soaking mother liquor could improve the ligand solubility in some cases (Hassell et al., 2007).Thus, the soaking condition was 90 mM Tris HCl pH 8.5, 35% Peg 3350, 10mM $MgCl_2$ (Table 4.6). Most of the inhibitors in this preliminary study were diluted from 20-50 mM stock in 90% v/v DMSO into the soaking condition to achieve 1-5 mM inhibitors for soaking. This step also diluted the DMSO down to 20% v/v in the soaking solution. The DENV 3 RdRp crystals could tolerates to 20-25% v/v DMSO without affecting crystals quality and diffraction. In the case of high aqueous solubility inhibitor or small ligand (with MW < 250 Da), 100-500 mM stock inhibitor can be prepared in 90-100% DMSO, which can further increase the final concentration for inhibitor (eg. 10 or 50 mM) while reduce DMSO to 10%

v/v in the soaking solution (Murray et al., 2007). This approach sometimes could increases the hit rate.

A substrate mimic and aqueous soluble inhibitor, 3'-deoxy-guanosine-5'-triphosphate (3'dGTP) was also used in soaking. The DENV 3 RdRp crystals were in the crystal growth mother liquor supplemented with 10 mM 3'dGTP for 20 h at 4 ºC (see material and methods). The result is discussed in Chapter 5.

In the case of protease I, co-crystallizations yielded very few protein-ligand complexes. Therefore soaking studies were devised by the authors (Hassell et al., 2007). The protease I crystal was fragile and most ligands were insoluble. The use of xylitol during the soaking stabilized the crystal and improved the solubility of ligands. Surprisingly, if the ligand was not mixed in the xylitol, the efficiency of obtaining protein ligand complexes greatly decreased. Therefore, it is important to have optimal soaking solutions to increase the hit rate of obtaining protein ligand-complex, as seen in the case of protease I. To investigate the DENV 3 RdRp protein-inhibitor complex studies in the future, we are now planning to employ 2D-NMR as a tool to screen the "real binder", a step that possibly improves the hit rate before proceedings with fragment based screening by X-ray crystallography. In addition, we have identified a stable DENV 3 RdRp construct, which has higher solubility in low NaCl concentration to fulfil the prerequisite condition for NMR screening (see appendix A2).

Fig 4.5.Typical damaged crystals.

DENV 3 RdRp crystals that were cracked either by intolerance to DMSO or inhibitor and thus cause the crystals loss of diffraction property.

4.4.4 RNA template

We set out to understand the interaction between RNA template and DENV RdRp, by soaking crystal into designed templates. We hope this will also serves as a model to build the RdRp-template-primer/substrate ternary complex. DENV 3 RdRp crystal were soaked into 0.2 M Tris HCl pH 8.5, 50 mM $MgCl_2$ and 30% w/v PEG 3350 that was supplemented with RNA templates, 5, 6 and 7 mers rU respectively (Table 4.6). The RNA template was dissolved in 50 mM Tris HCl pH 8.5 to make 5 mM stock. In RNA template soaking, the $MgCl_2$ concentration was reduced from 0.2 M to 50 mM for soaking studies to avoid RNA template precipitation. In the case of HCV RdRp template complex (PDB code 1NB7), no major conformational changes is observed, compared to the apo-enzyme. Furthermore, some RdRp/template complex models show that the RNA template tunnel of viral RdRp could accommodate 5-7mers RNA. Based on these observations, DENV 3 RdRp crystals were soaked with 1mM RNA 5-7 mer rU templates. Crystals soaked in rU_5 were either dissolved or cracked, indicating that the substrate may enter into crystals. Similar condition happened to crystals that soaked with rU_7 after 48 h or rU_6 after 72 h. Data diffraction for DENV 3 RdRp crystals that soaked in rU_6 for 24 h were collected. However, no extra electron density

peak was observed around the template tunnel. Further RNA template-DENV polymerase interaction study is needed to optimize the RNA template design and obtain the RdRp-template complex (length or sequence, see appendix A6). Co-crystallization or screening new crystal growth condition with natural viral RNA 3' UTR or the SLA sequence as template should increases the RNA template binding affinity to DENV RdRp and likelihood of obtaining the binary complex, as shown in the case of FMDV (PDB code 1WNE or HRV, PDB code 1TP7). The function viral RNA template and the 5' or 3' UTR was discussed in chapter 1. In addition, the RNA synthesis mediated by RdRp is also briefly outlined in chapter 5.

4.4.5 Divalent metal ion

DENV RdRp catalytic activity has been shown to be supported by divalent metal ions (Selisko et al., 2006). In addition, metal ion Mn^{2+} was also found to increase DENV polymerase catalytic activity in the presence of Mg^{2+}. We therefore selected Mn^{2+} ion for soaking to provide clues to understand the role of metal ion. The effect of these metal ions on RdRp structure is highlighted in chapter 5. The data collection and refinement statistics are listed in Table 4.3 and 4.4. Interestingly, the fingers subdomain of DENV 3 RdRp was disordered when crystal were soaked in 0.2M $MnCl_2$. Nevertheless, the palm and thumb domains were in good order and Mn^{2+} ions were shown clearly binding at the active site. By contrast, finger subdomains for crystal soaked in $MgCl_2$ are in good order. Thus, to avoid this, crystals were soaked in the presence of 10mM Mn^{2+}. Since higher Mn^{2+} concentration could lead to be inhibition of the polymerase, we suspected that the high Mn^{2+} concentration disordered the fingers subdomain and thus become detrimental to the polymerization activity.

4.5 Conclusion

In this work, we described the strategy used to obtain DENV 3 RdRp crystals that diffract to good resolution (1.85 Å). Firstly, small and initially poorly diffracting crystals that was not suitable for structure solution could be improved via an air dehydration procedure, resulting in a markedly improved diffraction pattern. The protein construct was found in complementation with cloning of different serotypes of the DENV catalytic domain RdRp (273-900), which discussed in chapter 3. Secondly, crystal quality could be further improved by addition of salts of metal ions and air dehydration treatment, leading to a high-resolution structure of DENV RdRp catalytic domain. In the second part of this work, X-ray crystallographic screening approach was employed to discover ligand for drug discovery. So far, no ligand (compound) was found to bind DENV 3 RdRp catalytic domain. However, the briefly discussed DENV3 RdRp crystal soaking study parameters could be used as a reference for the follow up of new ligand library in the future. To study the RdRp-RNA complex or RdRp-RNA-rNTP complex interaction, co-crystallization with the natural RNA viral template should increase the likelihood of obtaining the complex crystal. In the next chapter, the structural basis of DENV 3 RdRp is discussed.

CHAPTER 5

CRYSTAL STRUCTURE OF THE DENGUE VIRUS RNA DEPENDENT RNA POLYMERASE CATALYTIC DOMAIN

INTRODUCTION

The first breakthrough of polymerase crystal structure was that of Klenow fragment of *Escherichia coli* DNA polymerase I, solved in 1985 (Ollis et al., 1985). Subsequently, the structure of HIV RT was deciphered in 1992 (Kohlstaedt et al., 1992). The first atomic resolution of a RdRp was that of poliovirus 3D polymerase, a Picornaviridae polymerase (Hansen et al., 1997). Crystal structures of 13 RdRps from several virus families have been determined either as apo-enzymes or as complexes with incoming ribonucleoside triphosphates, primers, templates or small molecules inhibitors. These include RdRps which are members of the (1) Flaviviridae family such as DENV (Yap et al., 2007) and WNV (Malet et al., 2007), as well as BVDV (Choi et al., 2006) and HCV (Bressanelli et al., 2002; O'Farrell et al., 2003; Wang et al., 2003; O'Farrell et al., 2003; Wang et al., 2003), (2) Picornaviridae, including poliovirus (Thompson and Peersen, 2004), Foot and mouth disease virus (FMDV) (Ferrer-Orta et al., 2006b) and human rhinovirus (HRV) (Appleby et al., 2005) , and (3) Caliciviriadae such as rabbit hemorrhagic disease virus (RHDV) (Ng et al., 2002) and Norwalk virus (NV) (Ng et al., 2004). In addition, structures for RdRps of the double stranded RNA bacteriophage Φ6 and reovirus λ3 polymerase have also been reported (Butcher et al., 2001;Tao et al., 2002). Two RNA synthesis models are available for viral RdRp. The first model is deduced from dsRNA bacteriophage Φ6 RdRp (Butcher et al., 2001). This model represents the initiating RNA polymerization mechanism. The second model is derived from FMDV RdRp, which represents the elongative polymerization (Ferrer-Orta et al., 2007).

In Φ6 RdRp initiation complex, the nucleotide at the 3' end of the template is denoted as T1 and numbered sequentially from 3' to 5' while the daughter nucleotide is denoted as D1, D2 and so on. The base of T1 template threads into a specific binding pocket (S site) within the C-terminal domains (Fig 5.1). The C site is denoted for incoming NTP binding site for catalysis (C site). Interestingly, Mg^{2+} or Mn^{2+} ion binds at a site 6Å from the catalytic site (C site). The template T1 forms a Watson-crick base pair with D1 which is base stacked on Tyr-630, where is denoted as P site, a priming site that usually provided by residue either Trp or Tyr (P site). The structure reveals that the γ-phosphate of D1 remains intact during polymerization reaction and the tp moiety is in a distance to coordinate the divalent metal ions. The D2 that stacks on D1 and base pairs with the T2 of template, is positioned at C site with two Mg^{2+} ions coordinated by Asp which also at C site. The tp moiety is stabilized by Arg-270 and Arg-204 which may facilitate the NTP to shuttle between the site I and site C, and also responsible to guard the fidelity and stability of initiation complex. The substrate pore is denoted as site I (site I), which facilitates the interrogation of the template by incoming NTPs, is located 5Å from the NTP binding site at the active site (site C).

Based on the observation of the Φ6 RdRp initiation-complex, Φ6-template and Φ6-substrate complex, Butcher et al proposed a sequence of events that could result in the formation of the initiation complex (Fig 5.1). Step I-IV: the template enters the tunnel and interactions with the specificity pocket S site, NTPs occupy site I, presumably in rapid exchange. Step V: the D1 GTP binds to the initiation platform site P, stabilized by the hydrogen bonds of template and stacking interactions with Tyr-630. Step VI: the template ratchets back, facilitated by electrostatic attraction to Arg-268 and Arg-270, freeing T1 from the S site. Step VII: a second GTP D2 enters the P site to lock the initiation complex into active form, then catalysis of PPi release, freeing the residue Arg-268 and Arg-270 so that it can ratchet down and, Step

VIII: displacing the C-terminal domain of the protein. The template base changes orientation in response to the RNA oligonucleotides binding template. In the Φ6 RdRp, the T1 is rotated ~ 180°, T2 is stabilized by base stacking with T3. These changes results in the change of sugar conformation and side chains of interacting polymerase residues, leading to the extra hydrogen network that further stabilizes the RdRp-RNA-substrate complex, compared to DNA-polymerase complex.

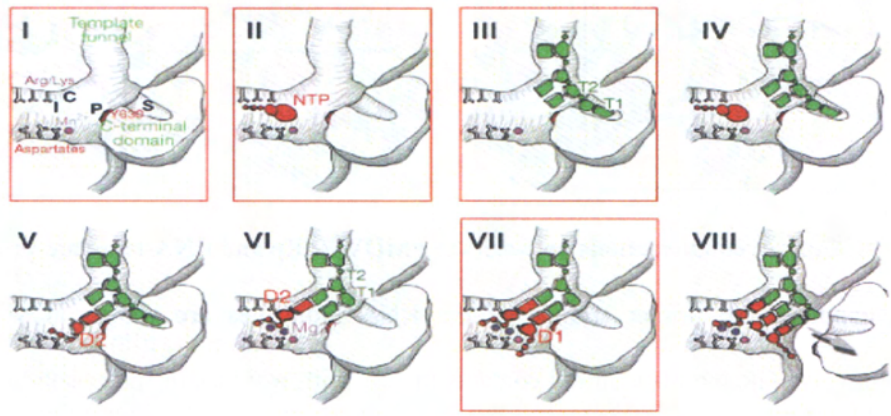


Fig 5.1.Model for initiation and chain elongation of the Φ6 RdRp.

The model illustrating key points in the reaction mechanism. Red boxes highlight experimental results. (I) Apostructure with bound Mn^{2+}. Binding sites are identified in black letters. (II) NTP bound in site I. (III) Template bound. (IV) Template bound and NTP non-productively bound bound at site I. (V) initial productive binding at site P. (VI) Template ratchets back. (VII) Second GTP bound at site P. Polymerization occurs. (VIII) Polymerization has occurred, releasing nascent duplex to ratchet forward, out of the active site.

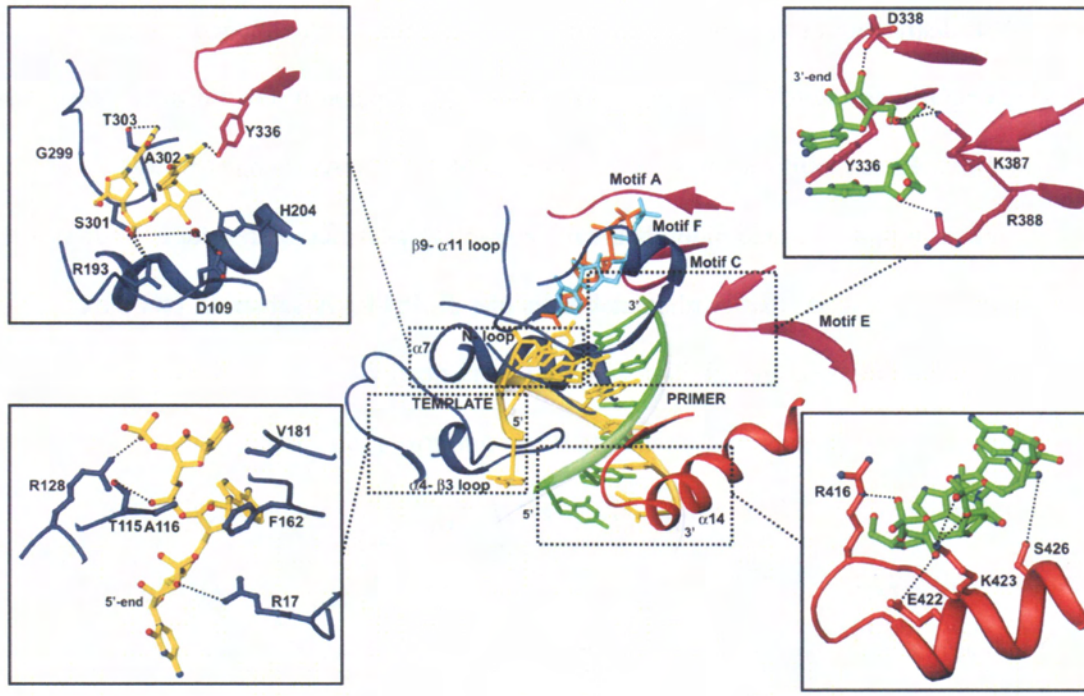


Fig 5.2. Conserved interactions between the FMDV RdRp and RNA-template-primers.

The template and primer strands of the RNA molecules are shown in yellow and green respectively. The template strand contacts mainly with residues in finger subdomain (blue). The 5' overhang region and the template binds the template channel, where the different residues of the N-terminal region and the loop α4-β3 of the polymerase drive the ssRNA to the active site (lower left). The template strand of the dsRNA product contacts different residues of helix α7 and the loop β9- α11 in its exit through the central cavity of the enzyme (upper left). The primer strand interacts with motif C and E of the palm domain, shown in magenta (upper right) and with α helix 14 of the thumb shown in red (lower right). Adapted from (Ferrer-Orta et al., 2007).

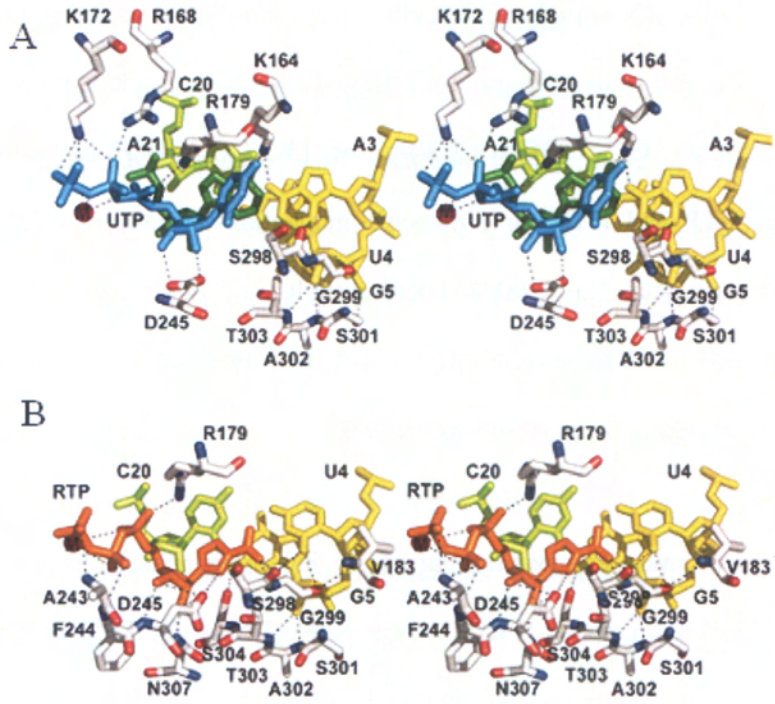


Fig 5.3. FMDV-RdRp complexes.

Non-conserved interactions established between the FMDV RdRp and its template-primer (TP)-RNA, (GCAUGGGCCC) and rNTP substrates. (A) enzyme-TP-ATP/UTP. (B) enzyme-TP-RTP. Metal ions are shown in red spheres, template and primer strand are in yellow and green respectively. UTP is shown in cyan and RTP is in orange. Enzyme-TP-ATP/UTP represents the ground binding state proposed by Cameron et al. The side chain of Asp-245 and Asn-307 occlude the ribose sugar binding pocket. By contrast, Asp-245 and Asn-307 in Enzyme-TP-reorientate to accommodate the ribose sugar and make interaction with the nucleotide. Adapted from (Ferrer-Orta et al., 2007).

Crystal structure of FMDV RdRp in complex with template primer RNA shows the RNA elongation process and the interactions between FMDV RdRp and the exiting duplex product (Ferrer-Orta et al., 2004). The structure reveals residues from fingers subdomains in direct

contact with the RNA, including residues from motif B, E in the palm and F in the fingers (Fig 5.2). Recently, the authors also reported the catalytic elongative complexes of FMDV RdRp-template (5'GCAUGGGCCC3')-primer (3' CCCGGG 5') in the presence either natural substrate (ATP and UTP) or mutagenic nucleotides (ribavirin, RTP and 5-fluorouridine, FUTP) (Ferrer-Orta et al., 2007). These complexes provide the model to define the critical residues involved in the successive replication events and insights for low fidelity of viral RNA polymerases. These complexes reveal:

(A) The recognition and positioning of the incoming nucleotide or analog that represents different replication events (Fig 5.3). The RdRp-RNA-ATP/UTP shows the translocation of the RNA product and the positioning of the new incoming UTP close to the active site while the RdRp-RNA-RTP complex shows the mutagen occupies the correct position at the active site for phosphoryl transfer. In addition, the RdRp-RNA-ATP and RdRp-RNA-FUTP respectively show the formation of new base pair and translocation of the RNA product. Comparison of the RdRp-RNA-ATP/UTP and RdRp-RNA-RTP identified the role of Asp-245 (motif A) and Asn-307 (motif B) in nucleotide recognition and correct positioning of the ribose sugar in the binding pocket (see 5.10). These key residues are structurally conserved both in DENV and $3D_{pol}$. In the FMDV UTP complex, the UTP is in the ground state binding which is proposed by Cameron et al., 2004 (Cameron et al., 2004) (see 5.10 for detail). The side chain position of Asp-245 and Asn-307 blocks the ribose binding. In addition, the tp moiety of UTP is also coordinated by residues from motif F at the nucleotide entry tunnel which is not accessible by active site. The observed role of Asp-245 and Asn-307 is consistent with the kinetic analysis on $3D_{pol}$, poliovirus (corresponding to Asp-538 and Asn-609 in DENV, Asp-238 and Asn-297 in $3D_{pol}$, see 5.10).

In the case of RTP complex, Asp-245 re-orientates its rotamer conformation and allows the positioning of the RTP sugar to be fit into ribose-binding pocket. This complex represents the second step of nucleotide binding of conformational change in $3D_{pol}$ which is proposed by Castro et al., 2005 (Fig. 5.3). Several interactions have been observed, compared to the ground state binding in UTP complex. The 2'-OH of RTP forms a double hydrogen bond with the side chains of Asn-307 and Asp-245. Furthermore, the 3'-OH forms a hydrogen bond with the N- atom of the main chain of Asp-245. In addition, the tp moiety of RTP is orientated to form extensive network with residues from motif A (Ala-243 and Phe-244), motif F (Arg-179) and one metal ion. The tp moiety conformation is further stabilized by the interaction between the β-phosphate and the 3'-OH of the sugar. The 3'-OH of the primer nucleotide is also positioned to form a hydrogen bond with the catalytic Asp-338. This residue also binds a metal ion that coordinates the tp moiety of the incoming NTP. These observations in the complexes are consistent with the kinetic study in $3D_{pol}$.

(B) The positioning of the acceptor base of the template strand in the complexes (Fig 5.2). The acceptor base is located adjacent to the nucleotide-binding site that accessible to the incoming nucleotide. Comparison of the four structures in the active site reveals common contacts mediated by residues of motifs C and E of the palm domain that stabilize the position of the 3' end of the primer strand in an orientation for RNA elongation (Fig 5.2). Motif A, B of the palm and motif F in fingers was also found to help the positioning of the acceptor base of the template strand and participate in the recognition and binding of the incoming rNTP (Fig 5.2). Interestingly, motif F position in DENV RdRp-RNA template binding is different from FMDV. This could probably be due to evolutionary pressure to distinguish specific protein-RNA template interaction. The position of Ser-298 and Thr-303 that are involved in template and incoming nucleotides interaction are critical in virus

replication. This was shown by mutagenesis (Ferrer-Orta et al., 2007). Interestingly, these residues are also structurally conserved in DENV RdRp (Ser-600 and Thr-605) and suggest the conserved role of these residues.

Objective of the study

In the previous two chapters, the strategies that used for DENV RdRp protein production, screening for crystallization conditions, good resolution crystal diffraction were reported. Since virus encoded polymerases present as highly druggable targets for the development of antivirals, we set out to understand the structure of DENV 3 RdRp and provide clues to understand its role and function. HIV-RT drug discovery have been greatly facilitated and advanced by the presence of wealth information on structural data. Similarly, we would like to understand (1) overall structure of DENV polymerase (2) the precise location of the nuclear localization signal (NLS) domain and how its folding affects polymerase enzymatic activity, (3) key amino acid residues involved in substrate binding and divalent metal ions coordination and (4) understand mechanisms of polymerization and fidelity for DENV polymerase. Such structural understanding in multifunctional DENV polymerase could eventually be used in facilitating rational drug design.

RESULTS AND DISCUSSION

5.1 Overall structure of DENV RdRp

The distribution of domains along the DENV3 NS5 sequence is schematically depicted in Fig 1A. The DENV 3 RdRp has overall dimensions of approximately 65 x 60 x 40 $Å^3$. Its architecture assumes the canonical right hand conformation consisting of fingers, palm and thumb, characteristic of known polymerase structures, but with the NLS region playing an important role in the formation of the observed structure (Fig 5.4). Interestingly, the NLS

domain signatures described previously (Brooks et al., 2002) are distributed between the fingers and thumb subdomains. Three helices (α2- α4) are incorporated within the thumb and the fingers (α 5- α7) subdomains respectively and the connection between them is realized through the segment linking helices α4 and α5. Truncations within the NLS region are thus likely to destabilize the protein structure, an observation which explains *a posteriori* the poor expression/solubility of constructs lacking part of this region. A total of 27 α-helices (α3 is a 3_{10} helix) and 7 β-strands are clearly defined in the current DENV3 RdRp model and an amino-acid sequence alignment of the RdRps from several flaviviruses including the four DENV serotypes is shown in Fig 5.5. The DENV 3 RdRp fold is analogous to the WNV polymerase (with a r.m.s. deviation of 1.23 Å for 313 superimposed Cα). The corresponding r.m.s. values for the three polymerase subdomains superimposed individually are 1.9, 0.8 and 1.0Å (for finger, palm and thumb subdomains respectively) suggesting that no gross conformational change occurs within the individual subdomains of the RdRps from WNV and DENV3. Their relative orientations, however, differ as displayed in Fig 5.6. Overall, the DENV 3 RdRp assumes a more "open" conformation compared to the conformation observed for WNV RdRp. In the latter, the fingers subdomain has rotated towards the thumb by an angle of approximately 8 ° (Fig 5.6).

5.2 The NLS region

The amino-terminal region comprising the b-NLS and ab-NLS sequences (residues 316-415) forms a peculiar feature of DENV RdRp, and such sequences are neither present in the HCV nor in the BVDV polymerase. The structure adopted by the b-NLS (residues 316-368) includes a helix-turn-helix motif that lies on the top of the thumb domain and helices α7 and α6 of the ab-NLS (residues 369-415) are buried in the fingers and between the fingers and palm domain respectively (Fig 5.4). Previous biochemical pull-down assays have shown that

the βNLS region of NS5 from DENV 3 interacts with the NS3 helicase, probably via a well conserved stretch of 20-amino acids. This segment (341- AMTDTTPFGQQRVFKEKVDT) within the βNLS (Brooks et al., 2002; Khromykh et al., 1998) forms part of a mobile region of the thumb subdomain. In the present structure, residues 350-360 which contain the 3_{10} helix α3 are properly oriented to interact with the importin β site with residues Arg-352, Phe-354, Glu-356 and Lys-357 positioned to interact with its binding partner. Together, the βNLS and the α/βNLS (Khromykh et al., 1998) form part of a functional import system of flaviviral NS5 to the nucleus of the host cell, where it contributes to a significant proportion (ca 20%) of the total RdRp activity from cells infected with West Nile virus, Japanese encephalitis virus (JEV), and DENV (Uchil et al., 2006). Furthermore, a link to the observed dengue pathogenesis is suggested by the nuclear location of NS5 in a report showing that the chemokine Interleukin-8 (IL-8) is induced after DENV 2 infection, presumably via a direct interaction with a CAAT/enhancer binding protein (Medin et al., 2005).

5.3 Fingers subdomain

The fingers subdomain comprises residues 273-315, 416-496 and 543-600. As shown by higher temperature factors, the fingers subdomain and especially the "fingertips region" appear quite mobile compared to the other palm and thumb subdomains. Residues 311-317, 408-415 and 454-466 (preceding motif F) were not observed in the electron density map and were thus not included in the present model. An SDS PAGE and western blot analysis of dissolved protein crystals ruled out proteolytic cleavage, indicating high mobility of these segments (data not shown).

In spite of a low sequence similarity, the extended "fingertip" feature is similar to other viral RdRps including primer-independent RdRps such as BVDV, HCV, and Φ6 (PDB codes 1S48,

1GX6, 1HHS respectively) (Bressanelli et al., 2002; Butcher et al., 2001; Choi et al., 2004; Khasnis and Nettleman, 2005; Gelpi et al., 2005) or primer-dependent RdRps FMDV and RHDV (PDB codes 1WNE and 1KHV respectively) (Appleby et al., 2005; Ferrer-Orta et al., 2004; Clark et al., 2005). The base of the fingers domain forms a concave surface shaped by the solvent-exposed residues of helices (α6, α14 and α15), near the N-terminus of the protein (Fig 5.4). This concave surface could be involved in accommodating the N-terminal MTase domain. A more precise mapping of the interaction between the MTase and the RdRp domains has been provided by Egloff et al for the WNV structure (Malet et al., 2007).

Compared to RdRps of HCV and BVDV, several differences can be noted in the DENV RdRp molecule. In DENV 3 RdRp, an additional N-terminal segment of 35 amino-acids (absent in HCV and BVDV) folds into helix α1 and strand β1, connecting to helix α2 of the NLS which is deeply buried in the thumb domain. Two flexible loops L1 and L2 (residues 309-321 and 342-347), linking the fingers and the thumb domains are likely to transmit conformational changes between the two domains and/or restrict their individual mobility. Furthermore the location of the template tunnel beneath these connecting loops suggests that any transition from a 'closed" to an "open" conformations of the RdRp during replication could be modulated by these loops (Ago et al., 1999; Khromykh et al., 2001b). The additional β1 strand belongs to a three-stranded anti parallel β-sheet that is surrounded by helices α1 and α5. The corresponding region in the HCV polymerase is occupied by a five stranded anti-parallel β sheet and such formation of β sheets is not observed in the BVDV RdRp. The L3 loop (residues 405-418) connecting α7 and α8 is highly mobile. This loop is also rightly positioned to regulate access of the ssRNA substrate at the entrance of the template tunnel. By contrast, the fingertip of the corresponding loop in BVDV and HCV twist away from the active site towards the fingers domain.

5.4 Palm domain and the catalytic active site

The palm domain consists of residues 497-542 and 601-705 and is composed of a small anti-parallel β-strand platform β4- β5, surrounded by eight helices α11- α13 and α16- α20 (Fig 5.4). Overall, the palm domain appears the most structurally conserved among all known polymerases, reflecting the preservation of the architecture of the catalytic site during evolution. Indeed the catalytic domain of DENV RdRp shows good superimposition with other RdRps (eg. BVDV or HCV), with r.m.s deviations of ~1.7Å and with the active site residues superimposing closely with the equivalent residues from other RdRps.

A comparative analysis of RNA polymerases from positive strand viruses have identified four out of six conserved sequence motifs responsible for NTP binding and catalysis, in the palm domain (Fig 5.4b) (Stiasny and Heinz, 2006; Nomaguchi et al., 2003).The GDD catalytic active site (motif C, comprising Asp-663 and Asp-664) is located in the turn between strands β4 and β5. The anti-parallel β-sheet that houses the active site has shorter strands of approximately 10Å (like in WNV RdRp), compared to the equivalent β-strands found in BVDV and HCV, which are about 20 Å long. In addition, the connection between motifs B and C appear more elaborate in RdRps from flaviviruses (comprising helices α17 and α18) as compared to RdRps from other genera which only have a short α helix directly attached to the β sheet of the active site. These differences constitute interesting signatures differentiating RdRps of flavivirus from pestivirus and hepacivirus that occurred during the divergent evolution of these viruses. As in the case of HCV, the active site of the RdRp from DENV 3 is encircled by several loops that contribute to shape the tunnel into which the template RNA can gain access to the catalytic site. The metal binding is discussed elsewhere in this chapter (see section 5.8.2).

5.5 Thumb domain

The thumb domain (residues 706-900), which forms the C terminal end of the RdRp of DENV, is the structurally most variable among known polymerase structures. It contains two conserved sequence motifs. Motif E forms an anti-parallel β sheet wedged between the palm domain and several α helices of the thumb domain. Interestingly, the loop (L1) that links helices α21 and α22 in RdRp of DENV 3 (residues 740-747) adopts a different conformation compared to HCV and BVDV where it forms a β-hairpin (Fig 5.4a). Together with the fingertips, the path adopted by this L1 loop contributes to shape the RNA template tunnel. A second loop spanning amino-acids 782-809 forms the priming loop that partially occludes the active site (Fig 5.4a & b). The path followed by this long and potentially flexible loop is clearly defined in the electron density map. It is stabilized by several "internal" (intra-loop) interactions including hydrogen bonds formed between Thr-794 and Ser-796, a salt bridge between Glu-807 and Arg-815 that projects from helix α24, as well as stacking interactions between the δ-guanido group of Arg-749 and the indole ring of Trp-787. In addition, hydrogen bonds formed by residues from this loop with residues projecting from helix α21 contribute to maintain its orientation with respect to the protein. Together with loop L3, the priming loop forms the upper part of the RNA template tunnel and is thus likely to regulate access and exit into the active site (Fig 5.4). Replacement of the corresponding β-hairpin with a short turn in the RdRp from HCV converts the protein from a solely primer-independent enzyme to a primer-dependent one, which can use double-stranded RNA as substrate (Hong et al., 2001). Of note, RdRps from HCV, NV or Φ6 also possess unique C-terminal region that fold back into the active site cleft (Adachi et al., 2002; Ng et al., 2004). Collectively, these specific structural features differentiate primer independent RdRps (eg: BVDV, HCV, DENV) from primer-dependent RdRps (eg: FMDV, RHDV). Accordingly, RdRps of poliovirus and calicivirus which lack such features initiate RNA transcription using a

template/primer duplex as a substrate for polynucleotide incorporation (Ng et al., 2002). In the DENV RdRp, the distance between the last residue (Met-883) that is visible in the electron density and the GDD motif C is 39 Å, thus ruling out a role played by C-terminal residues in regulating access to the active site.

5.6 Zinc binding sites

Unexpectedly, our crystal structure reveals two zinc binding pockets, Zn1 and Zn2, in the thumb and fingers subdomains respectively (Fig 5.4d). Since no zinc ions were included in the crystallization conditions nor during protein purification, they are likely originated from the bacterial growth media. One zinc atom (Zn2) is coordinated by His-712, His-714, Cys-728 of motif E and Cys-847 of helix α26. A strong peak (at 18σ) was observed in the residual electron density map and this site seems fully occupied in the protein with a low temperature factor of 24 $Å^2$ (Fig 5.4). The residues involved in coordinating the zinc ion Zn2 are conserved in the four serotypes DENV 1-4 and also in the RdRp from YFV. Residue His-714 is substituted by a Threonine in the RdRp from WNV and a disulfide bridge linking the two evolutionary conserved Cysteine residues is observed instead (PDB code 2HFZ). The second structural zinc (Zn1) located in the fingers subdomains is coordinated by Cys-446, Cys-449 and His-441 and is conserved in the WNV RdRp molecule. Interestingly, the HIV RT assumes a sharp turn after motif E that serves as a pivot point for movements of the thumb subdomain, upon template/primer binding (Jager et al., 1994). The zinc ion Zn2 likely contributes to the structural stability of the region near motif E of the DENV polymerase. Interestingly, this pocket is also located near the functionally important residues Ser-710 and Arg-729 which bind to the incoming rNTP. The Zn2 ion could therefore also play a role in regulating conformational switches within the thumb subdomain, that occur along the

reaction pathway. Further mutagenesis experiments are needed to address the exact functional significance –if any- of these two zinc binding sites.

5.7 The Priming loop and its role in the initiation complex

A high concentration of GTP is required for de novo initiation by the DENV RdRp, regardless of the precise nucleotide sequence at the 3'end of the RNA template (Nomaguchi et al., 2003). The structure of the complex between DENV RdRp with the 3'dGTP nucleoside analog should shed light on the GTP binding site(s) and its effect on de novo initiation. No large conformational change of DENV RdRp occurs upon binding 3'dGTP. The electron density map obtained from a crystal soaked with 3'dGTP and Mn^{2+} revealed clear extra density corresponding to the triphosphate (tP) moieties of 3'dGTP near the priming loop about 7Å away from the catalytic site (Fig 5.4e). Only very weak electron density was present for the guanosine base and the sugar moiety indicating their mobility in the absence of a template. No metal ion was observed to coordinate the tP moiety.

Comparisons with the RdRps from BVDV and HCV in complex with nucleosides reveal the presence of an rNTP substrate positioned similarly in the vicinity of the priming loop (P site). In the DENV 3 RdRp molecule, the tP moiety of 3'dGTP is coordinated by Ser-710, Arg-729, and Arg-737 (Fig 5.4f). These residues are strictly conserved across positive strand RNA viruses known to initiate replication using a de novo mechanism (Fig 5.5), including DENV1-4, YFV, JEV and WNV. In several de novo RdRps initiation complexes reported, the base of the ribonucleotide bound at the P site stacks against a tyrosine residue (Tyr-630 for the Φ6 bacteriophage, Tyr-448 for HCV and Tyr-581 for BVDV) (Butcher et al., 2001; Choi et al., 2004; Khasnis and Nettleman, 2005; Gelpi et al., 2005). The orientation of the tP moiety in our crystal structure (which was inferred from the superimpositions described

above) places the base at a right distance to form stacking interactions with Trp-795. Additional interactions with the 3'dGTP α-phosphate include hydrogen bonds with Thr-794 and Ser-796. Residue Arg-737, projecting from helix α21, makes a salt bridge with the β phosphate and Arg-729 (motif E) with the γ phosphate. Residue Ser-710 (from motif E) makes a hydrogen bond with the γ phosphate. Those residues thus provide the platform for de novo RNA initiation by the flavivirus polymerase domain. In a separate study, residues Ser-498 and Arg-517 in BVDV (corresponding to Ser-710 and Arg-729 in DENV RdRp) were mutated individually to an alanine, in order to test their contribution to an elongative (primer-dependent) versus de novo (primer-independent) mode of RNA synthesis (Lai et al., 1999). The de novo mode of RNA synthesis was almost completely abolished by these single mutations, whilst RNA elongation was reduced only by a factor of approximately 2 to 9 fold. Thus, a mutation of these residues from the BVDV RdRp was able to confer specificity for a primer-dependent mechanism as opposed to a de novo mechanism of RNA synthesis. The structural conservation of these residues within motif E with the BVDV RdRp suggests a similar phenotype when Ser-710 and Arg-729 are mutated in the DENV RdRp enzyme. A comparable study of the HCV RdRp demonstrated a similar role played by Arg-386 and Arg-394 (also leading to a severe decrease in de novo initiation). These two residues are structurally equivalent to Arg-729 and Arg-737 in the DENV 3 RdRp (Ranjith-Kumar et al., 2003). Taken together, these data suggest an essential role for this GTP-binding site for de novo initiation of RNA synthesis by the DENV RdRp. Interestingly, one Zinc atom (Zn2) located between the anti parallel β-sheet β6-β7 and helix α26 (Fig 5.4d) is found in the vicinity of Arg-729 and could thus play a role in modulating de novo initiation.

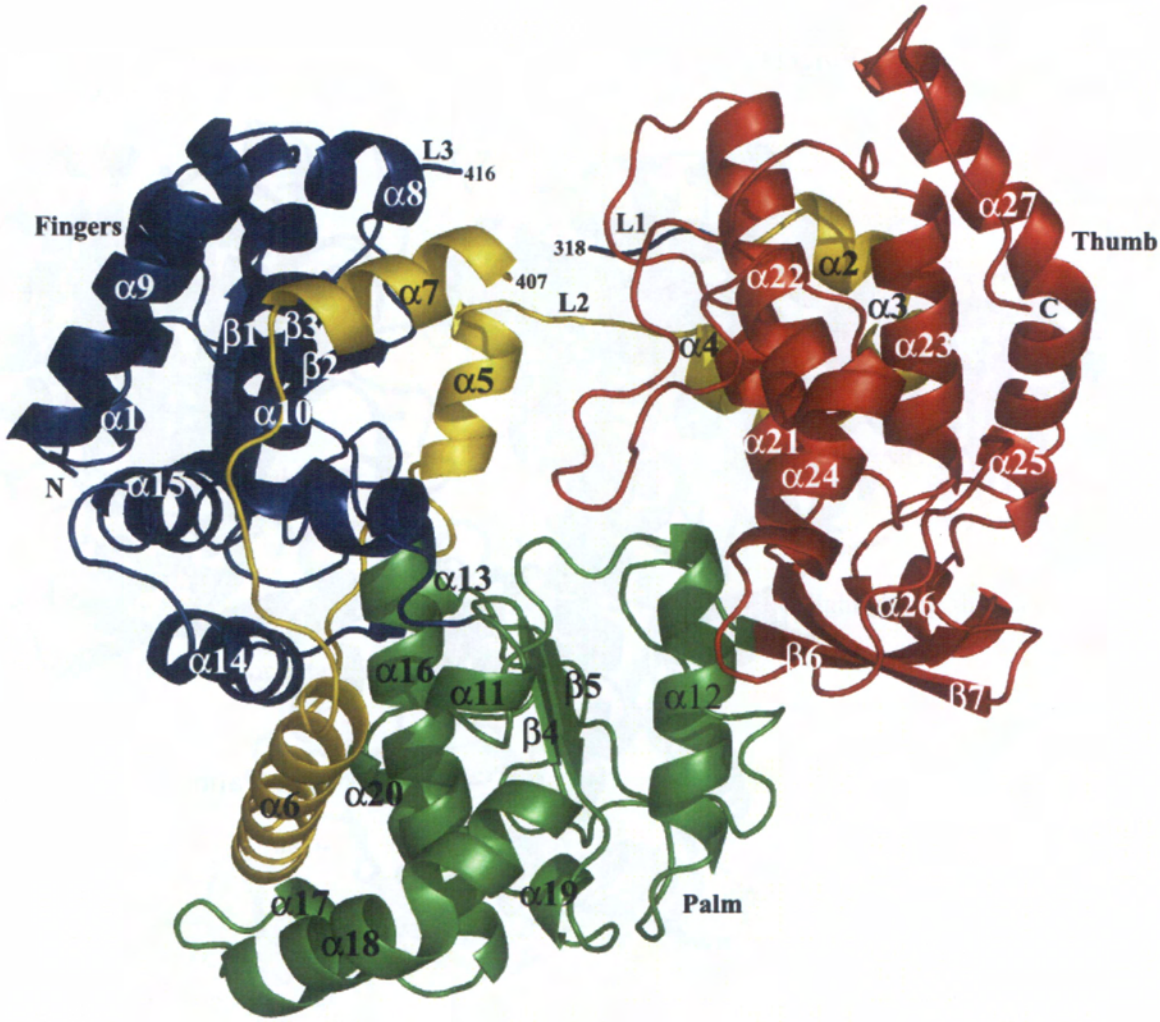
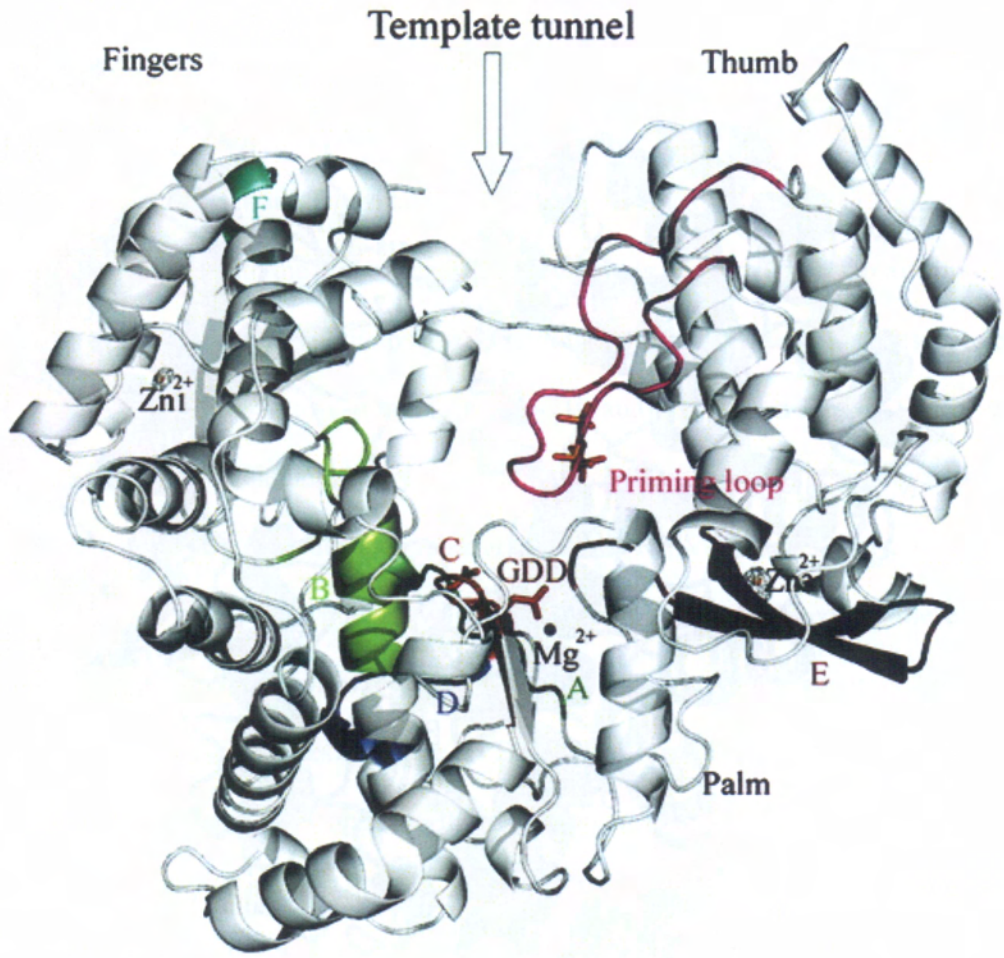


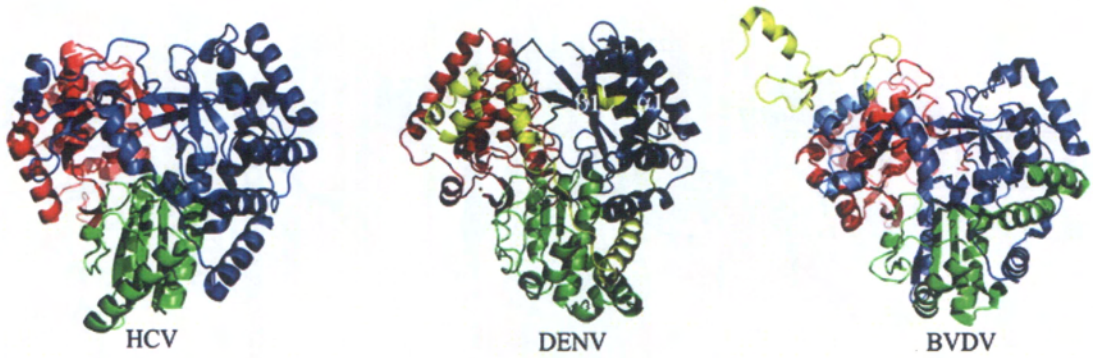
Fig 5.4.Structure representation of DENV 3 RdRp.

(A) Ribbon representation of the overall structure of DENV 3 RdRp (residues 273-900). The various subdomains are colored as follows: Fingers (blue), NLS (yellow), Palm (green), and Thumb (red). Secondary structure elements as defined in Fig 3 are labeled. The N- and C-termini are displayed and the residues at the boundaries of the missing regions in the model are numbered.

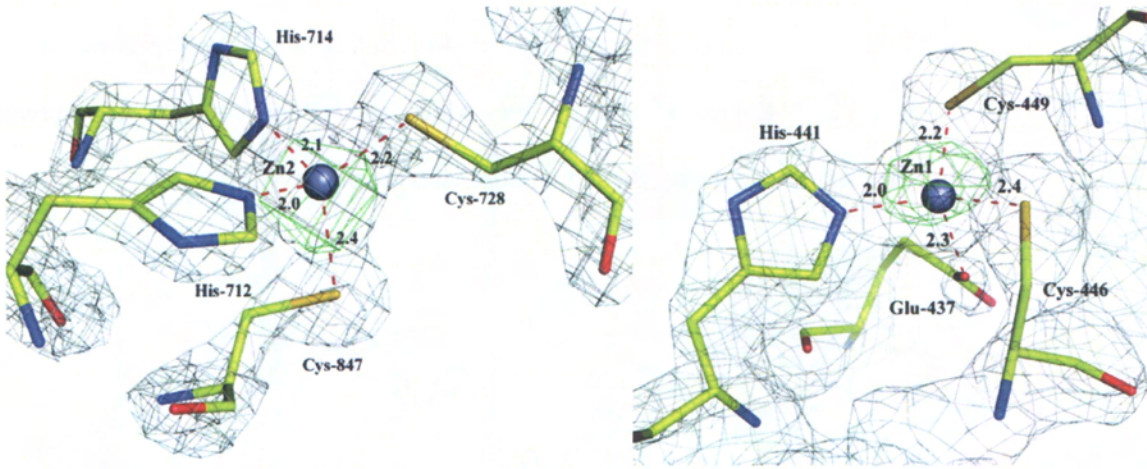


(B) The six conserved motifs present in various RdRps structures are mapped onto a ribbon representation of the DENV 3 RdRp catalytic domain and colored as follows: motif A (dark green), motif B (green), motif C (red), motif D (dark blue), motif E (magenta) , motif F (light blue) and priming loop (pink). Asp-663 and Asp-664 from the motif C of the catalytic site are shown in sticks representation and labeled. The divalent metal ions Zn^{2+} and Mg^{2+} are represented by spheres colored in gold and purple respectively. The inset depicts the interactions established by the triphosphate moiety of the 3'dGTP inhibitor. The final difference electron density map with Fourier coefficients F_o-F_c, where the atoms from the inhibitor have been omitted from the calculation is contoured at 3.0 σ. Residues Ser-710, Arg-729, Arg-737, Thr-794, Trp-795 and Ser-

796,which are making contacts are displayed as sticks and the distances with the α-, β- and γ-phosphates are displayed.

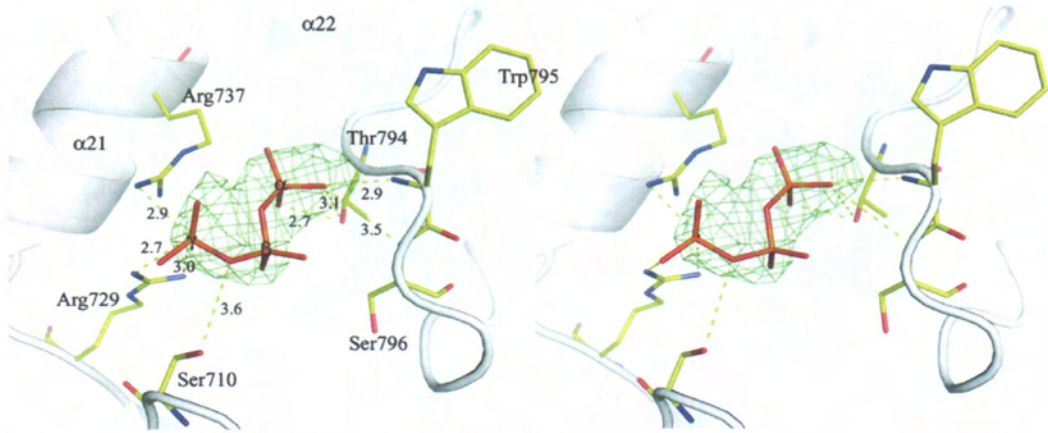


(C). Side by side views of the DENV, HCV, and BVDV RdRps, highlighting their structural differences (see text for details).

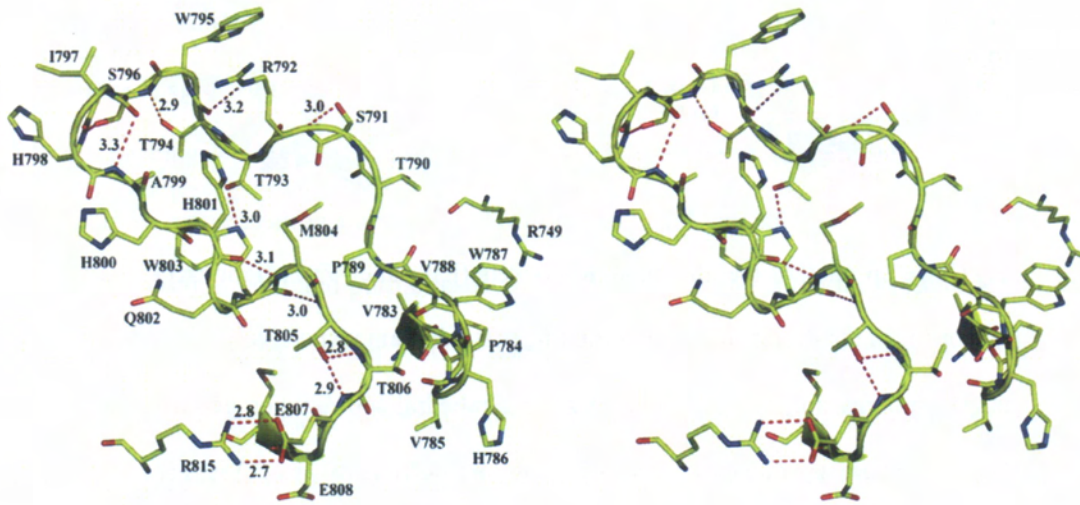


(D) Close up view of the zinc binding pockets (left panel, Zn2; right Zn1) located in the vicinity of motif E in the thumb and within the finger subdomain respectively. The zinc ions are displayed as gray spheres, and residues providing the tetrahedral coordination geometry are shown as sticks. The refined electron density map (with Fourier coefficients $2F_o$-F_c) is displayed at a level of 1σ.

The anomalous Fourier maps using phases from the refined final model are contoured at a level of 5 σ and displayed in green.



(E) Stereoviews of the interactions established by the tp moiety of the 3'dGTP inhibitor. The final difference electron density map with Fourier coefficients F_o-F_c, where the atoms from the inhibitor have been omitted from the calculation, was contoured at 3.0σ. Residues Ser-710, Arg-729, Arg-737, Thr-794, Trp-795 and Ser-796, which are making contacts, are represented as sticks, and the distances to the α-, β-, and γ-phosphates are displayed.



(F) Stereoviews of the interactions of the priming loop (see text for details)

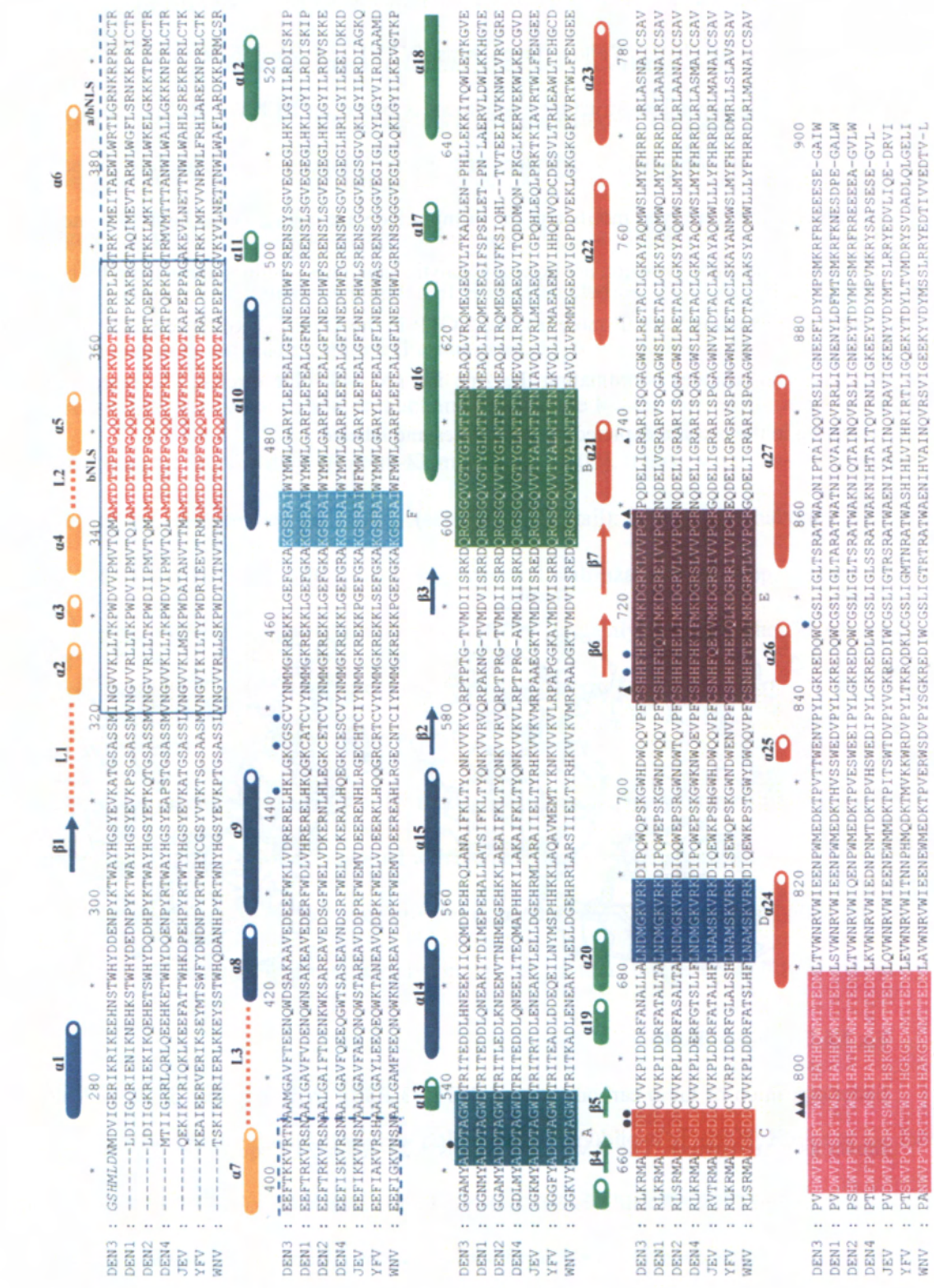


Fig 5.5. Structural based sequence alignment of the flavivirus RdRp domains

Structure based sequence alignment of the flavivirus RdRp domains from DENV 1-4, Japanese encephalitis virus (JEV), Yellow fever virus (YFV) and West Nile virus (WNV). Residues numbering and secondary structure assignment is based on DENV 3 RdRp structure; α-helices and β-sheets are depicted as cylinders and arrows, respectively. The fingers, palm, thumb subdomain, the conserved motifs and the priming loop are colored as in figure 5A and figure 5B. The bNLS and a/bNLS region are boxed. Catalytic aspartates residues are marked by black circles. Residues interacting with tP moieties of 3'dGTP in the DENV 3 RdRp complex are marked by triangles and residues coordinating the zinc ion are with blue circles. The three loops L1-L3 discussed in the text are depicted as dotted lines and labeled.

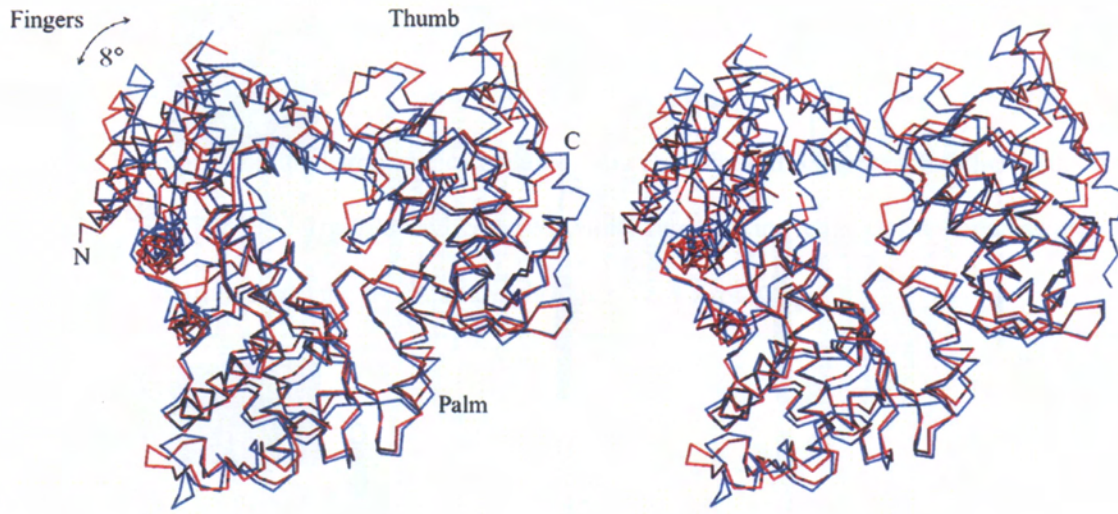


Fig 5.6.Cα trace of RdRp.

Superimposition of the α-carbon trace of DENV 3 RdRp (red) with that of WNV RdRp (blue), shown as a stereoview. The RMS deviation is 1.23 Å for 313 equivalent α-carbon atoms. The overall fold of DENV 3 RdRp is well preserved compared with that of WNV RdRp. Rotation of the fingers at an angle of ~8° is observed between the two structures (see the text for details).

5.8 Metal ion binding

Polymerases are metal-activated enzymes that use divalent metal ions for nucleotide polymerization. The Mg^{2+} ions coordinate the nucleotides, catalyze the formation of the phosphodiester bond, and transiently stabilize the by-product PPi (Joyce and Steitz, 1995; Salgado et al., 2004). The inclusion of 1-2 mM Mn^{2+} in the reaction mixture that contains Mg^{2+} is known to stimulate a number of RdRps (including HCV and DENV) and higher concentrations could lead inhibition (Ferrari et al., 1999). In addition, Mn^{2+} also significantly increases the de novo initiation product and reduces the K_m for the initiation nucleotide GTP in the case of HCV RdRp (Ranjith-Kumar et al., 2002). Thus, Mn^{2+} was proposed to be essential for RNA replication fidelity (see 5.10 for detail). A separate study also shows that both Mn^{2+} and Mg^{2+} are critical to HCV structural stabilization and support conformational changes to promote favorable active site geometry for catalysis (Benzaghou et al., 2004). In the case of DENV 2 FLNS5, we found that the use of 1-2 mM Mn^{2+} in the inclusion of reaction mixture (contains 2 mM Mg^{2+}) was the optimal assays buffer. Without it, the polymerase reaction was significantly reduced. However, Mg^{2+} and Mn^{2+} ions (concentration ranged from 1 to 10 mM) were found to have no effect on the melting temperature for DENV 3 RdRp and DENV 2 FLNS5 respectively, using thermoflour experiments (data not shown). Therefore, these metal ions may no have impact on protein structural stability in this study.

5.8.1 Metal ion catalysis mechanism

Based on the structural data of several polymerases (Doublie et al., 1998; Kiefer et al., 1998), Steitz proposed a "two metal ion" mechanism for polymerase- catalyzed nucleotidyl transfer reactions. In the structure of T7 DNA polymerase, metal ion A was found to coordinate the α-phosphate, increases the nucleophilicity of the primer 3'-OH by lowering its pKa (position A) and facilitating the $3'O^-$ attack on the α-phosphate, as shown in Fig 5.7. Metal ion B that

contacts the γ and β-phosphate of the incoming nucleotide is likely to stabilize the oxyanion that forms in the penta-covalent transition state geometry and to facilitate the release of pyrophosphate (position B) (Fig 5.7). Steitz proposed that this mechanism is universal to RNA or DNA polymerase.

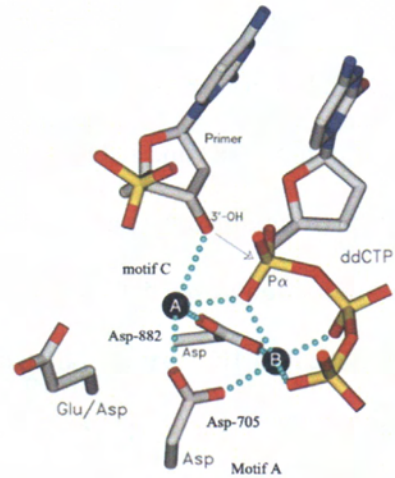


Fig 5.7.Polymerase catalyzed nucleotidyl transfer with two metal ions mechanism.
The divalent cation (Mg^{2+}, metal B, position B) is coordinated by the β- and γ-phosphates of the nucleotides, Asp residue located in motif A of all polymerases and likely water molecules. During catalysis, metal B orients the triphosphate in the active site and may contribute to charge neutralization. Once the nucleotide is in place, the metal A binds (position A). The metal A is coordinated by the 3'-OH, the α-phosphate and residue of Asp of motif A and C. Metal A lowers the pk_a of the primer 3'-OH to regulate the catalysis at physiological pH. The residues are refered to T7 polymerase in this figure. Adapted from (Steitz, 1998).

Recent studies on poliovirus RdRp ($3D_{pol}$) revealed that 2 protons transfer occurs in the nucleotidyl transfer transition state (Castro et al., 2007). The study provides evidence that a Lys-359 of $3D_{pol}$ serves as general acid for protonation of the ppi while Asp-328 and Asp-329

serve as general base to play role in primer 3'-OH deprotonation in polymerase-catalyzed nucleotidyl-transfer reactions. During the nucleotidyl transfer reaction, the proton from the primer 3'-OH is removed and donated to the leaving ppi (Castro et al., 2007). Lys-689 of DENV 3 RdRp is located to a position similar to the general acid Lys-359 of $3D_{pol}$. The proposed general base is catalytic conserved Asp residues (Asp-663 and Asp-664 in DENV RdRp). In a different study, Castro et al., 2005 also proposed nucleotide binding incorporation to RdRp is driven by the metal-liganded triphosphate (tp) moiety of the nucleotide at the active site (Castro et al., 2005). A conformational change occurs to bring the metal liganded tp moiety into appropriate position to interact with the conserved aspartyl groups of the enzyme. This state makes the active site ready to accept a second metal ion for catalysis to occur (see 5.10 for detail) (Castro et al., 2005).

5.8.2 Metal ion at active site of DENV RdRp

The RdRp structures of HCV, Φ6 and RHDV reveal that Mn^{2+} can either bind at the catalytic pocket of active site (Bressanelli et al., 2002) (Ng et al., 2002) or a specific allosteric position (Butcher et al., 2001). In order to study the Mn^{2+} binding, DENV RdRp crystals soaking were conducted (chapter 4). The DENV RdRp-Mn^{2+} complex shows the Mn^{2+}(I) ion is coordinated by Asp-663-664 of the catalytic site with ~2.5 Å, as shown in Fig 5.8. Based on the Φ6 RdRp initiation complex (1HHS), the rGTPs of Φ6 were superimposed and modeled in DENV RdRp. The model displays that Mn^{2+}(I) is also possible to interact with the primer 3'-OH and α-phosphate with ~1 Å and 2.6 Å respectively. The Mn^{2+}(I) site is located to a site similar to that of metal A which is proposed by Steitz (Fig 5.7). Compared to T7 polymerase, Asp-663 (motif C) and Asp-533 (motif A) of DENV RdRp-Mn^{2+} is similar to that of Asp-882 and Asp-705 of T7 polymerase respectively (Fig 5.7). Both Asp residues of T7 polymerase regulates metal A and B (Fig 5.7). In contrast, we identified Asp-664 of DENV RdRp

regulates two metal ions (position A and an allosteric site, see later) and no metal binding at position B (Fig 5.8). Interestingly, it is only in Φ6 RdRp initiation complex that the "two metal ions" mechanism in viral RdRp has been shown (Fig 5.8). Whereas the FMDV active elongation complex (2EFZ) shows the metal ion binds at position B and allosteric position respectively. Both complexes provide important understanding on different stage of viral replication. The difference between the absence or presence of extra metal ion respectively in FMDV position B and Φ6 allosteric pocket is an issue that needs to be resolved. The Mn^{2+}(II) of DENV RdRp is located at the pocket similar to allosteric position that is ~ 6 Å from the catalytic pocket in Φ6 RdRp (Fig 5.8). This Mn^{2+}(II) pocket is only unique in viral RdRp and observed in RdRp of Φ6 (IHHS, initiation complex), RHDV (1WNE), FMDV (2EFZ, elongation complex) and WNV (2HFZ). At this allosteric site, Mn^{2+} and Lu^{3+} was identified in RdRp of Φ6 and RHDV respectively while Mg^{2+} was found in the case of WNV and FMDV RdRp. Suprisingly, based on structural analysis on the Asp residues (motif A and C) in Φ6 RdRp and FMDV complex, we found that these residues are not in optimal geometry (>3 Å) for regulating the metal ions. This is in contrast with the tight metal ions-Asp catalytic residues interaction (< 2 Å) in HIV-RT-template-primer-substrate complex.

Compared to DENV RdRp-Mn^{2+}, crystals that were soaked with 0.2 M $MgCl_2$ only reveal single hydrated Mg^{2+} ion that is hexa-coordinated by Asp-533 (motif A) via a water molecule and by Asp-664 (motif C) at Mn^{2+}(II) binding site (Fig 5.9). Interestingly, higher Mg^{2+} concentration (0.2 M $MgCl_2$) and longer Mg^{2+} soaking time (20 h) was carried out and no Mg^{2+} is identified at the catalytic site of DENV RdRp (compared to 10 mM $MnCl_2$ and 12 h). This indicates that the occupancy for Mn^{2+} ion is higher than that of Mg^{2+} ion at the catalytic active site. Similarly, the occupancy for Mn^{2+} is also higher than that of Mg^{2+} in the HCV

RdRp catalytic site (Bressanelli et al., 2002). This could support the importance of Mn^{2+} ions inclusion in the polymerization assay.

In addition, Asp-533 of motif A DENV RdRp (Mn^{2+} and Mg^{2+}) was found not to interact with any metal ions, unlike in the proposed mechanism by Steitz. Instead, this residue coordinates metal ion (Mn^{2+}II or Mg^{2+}) via water molecules. In the case of WNV RdRp, Asp-536 and Asp-669 coordinate a Ca^{2+} ion (corresponding to Asp-533 and Asp-664 in DENV RdRp) (Malet et al., 2007). This Ca^{2+} ion is located to a site similar to Mn^{2+} (II) and ~ 3 Å further towards Asp-536 of WNV RdRp (Asp-533 in DENV RdRp). Mutation of this residue in BVDV (Asp-533 in DENV RdRp) abolishes the elongative synthesis and significantly reduces the de novo synthesis (Lai et al., 1999). This indicates the conserved role of this residue in RNA initiation and elongation process.

5.8.3 Conformational change of DENV RdRp

Structural comparison between RdRp-Mn^{2+} (brown), RdRp-Mg^{2+} (green) and RdRp without metal ion (blue, in complex with rGTP) did not show any major conformational change (Fig 5.9), except for β2-β3 of fingers subdomain, which is flexible and mobile. The β2- β3 of fingers subdomain is believed to regulate template binding during polymerization. Interestingly, the β2- β3 of fingers subdomain that belongs to Mn^{2+}, Mg^{2+} and no metal ion-RdRp structures respectively is observed to move sequentially from "close to open" form. However, it is unsure that the "closed" form of fingers subdomain for RdRp-Mn^{2+} is due to the effect of Mn^{2+} ion binding.

In addition, the *B*-factor for α1 and β1 of DENV RdRp-Mn^{2+} is higher than that of other two complexes. This observation could also explain the disorder of fingers subdomain when

crystals are soaked in high Mn^{2+} concentration (0.2 M $MnCl_2$, but not in the case of 0.2 M $MgCl_2$, see chapter 4). In summary, we need to construct a ternary complex structure (in the presence of template-primer/substrate) in order to investigate the structural basis of metal ion-substrate-protein interaction.

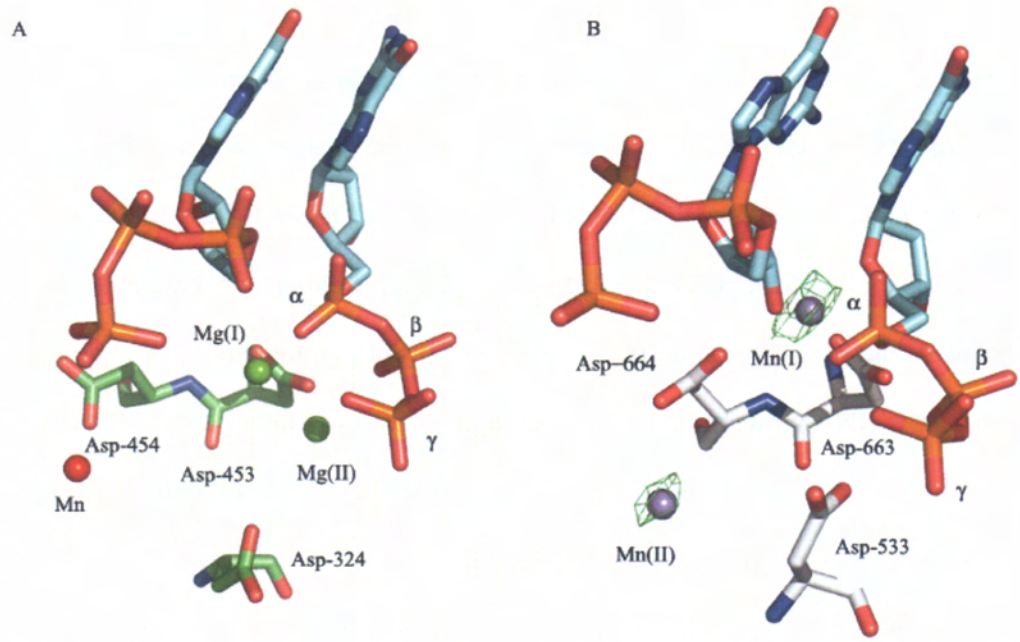


Fig 5.8.Metal ion binding site at the catalytic pocket and allosteric site.

(A) Φ6 RdRp initiation complex shows the catalytic active site with 2 rGTP. The Mg^{2+} is shown in green ball and Mn^{2+} is in red ball. The Mg(I) is in interaction with α-phosphate and 3'-OH of the first rGTP (left). Whereas the Mg(II) interacts with Asp-453, β and γ-phosphate. The Mn^{2+} is coordinated by Asp-454 and Asp-324 of motif A is too far to coordinate any metal ions. This model shows the Mg(I) and Mg(II) positions are in good agreement with the proposed "two metal ions" mechanism. (B) Mn^{2+} binding site for DENV RdRp-Mn^{2+} complex. The positions of 2 rGTP are modeled based on Φ6 RdRp. The Mn(I) is coordinated by Asp-663 (motif C) and water molecules. This Mn(I) is also in contact with the modeled 3'-OH rGTP (left) α-phosphate (right). While Mn^{2+} (II) is coordinated by Asp-664 (motif C) and water. The anomalous Fourier maps

using phases from the refined final model are contoured at a level of 4σ and displayed in green. Mn^{2+} is represented as a ball.

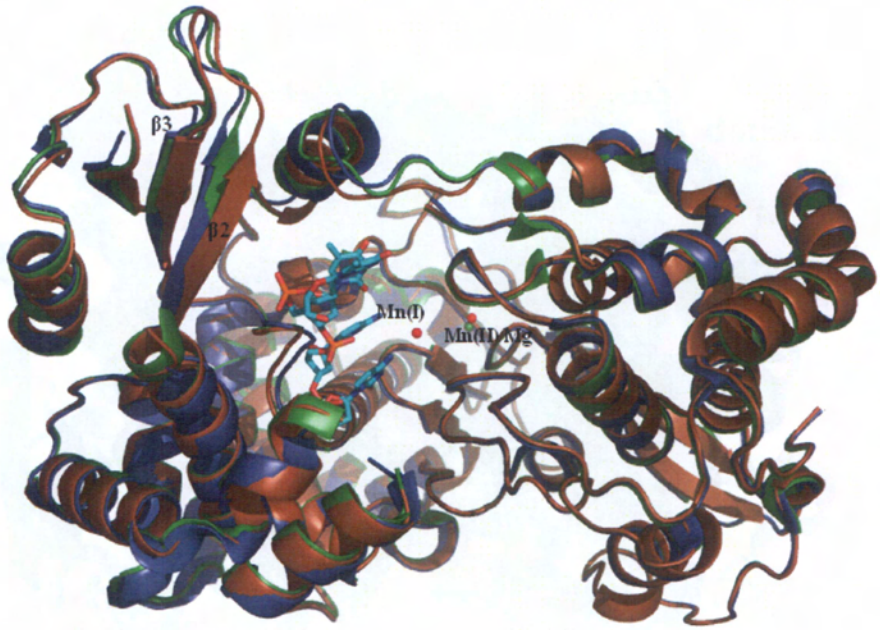


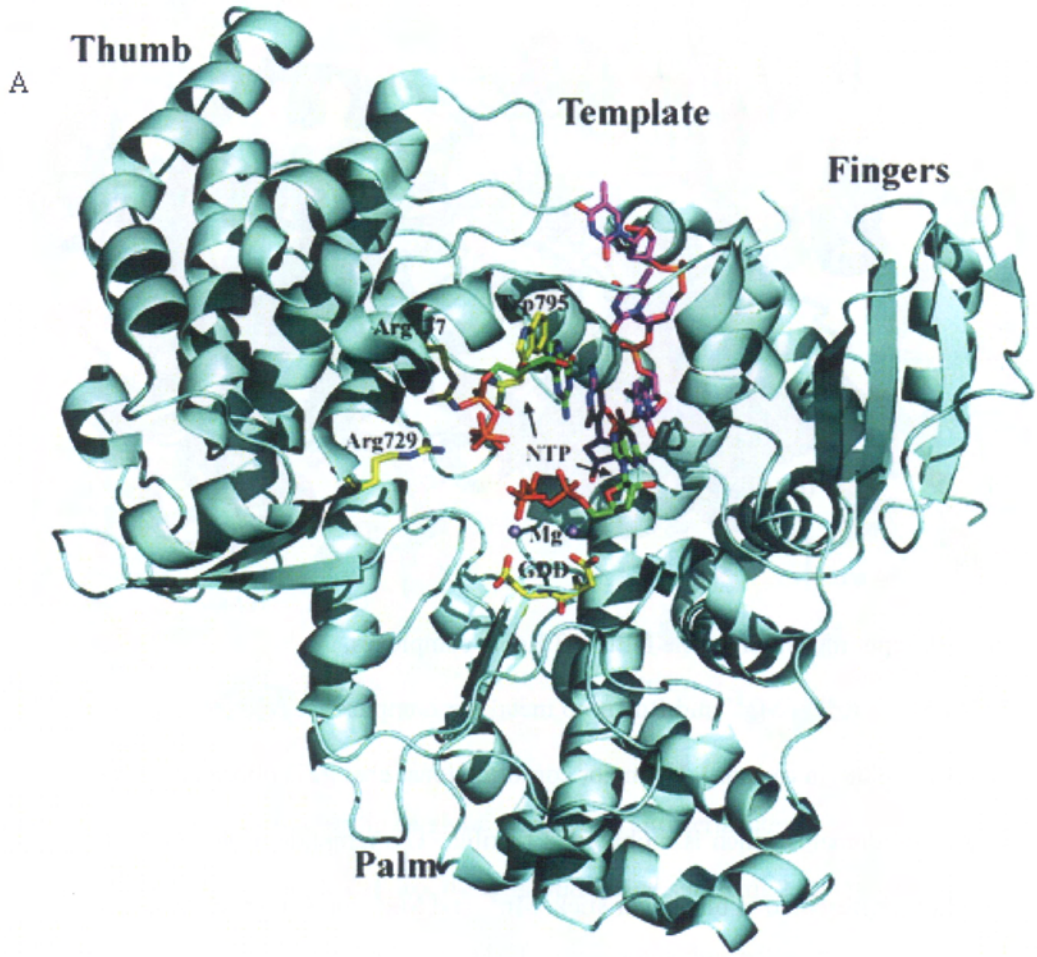
Fig 5.9.Superimposition of the DENV 3 RdRp complexes.

RdRp-Mn^{2+}, RdRp-Mg^{2+} and RdRp-no metal ion complex are respectively represented in brown, green and blue (in cartoon). No major conformational change is observed, except for the β2- β3 of fingers subdomain, which is mobile and flexible. The template of Φ6 RdRp is modeled into the structure, represented in blue color stick. Mn^{2+} and Mg^{2+} ion is respectively represented in red and green in ball form. Mn^{2+}(I) binds at catalytic pocket while Mn^{2+} (II) and Mg^{2+} share a similar allosteric site that also found in Φ6 RdRp, which is ~ 6 Å from the catalytic pocket.

5.9 A model for RNA synthesis

We built a model for a ssRNA template bound to the DENV 3 RdRp (Fig 5.10), based on the experimental complex observed between the RdRp of bacteriophage Φ6 and a template RNA strand (PDB code 1HI0). The ensuing position and orientation of the Φ6 RNA template

strand introduces minimal steric clashes with DENV 3 RdRp, suggesting that the mode of binding could be essentially preserved between the two enzymes.



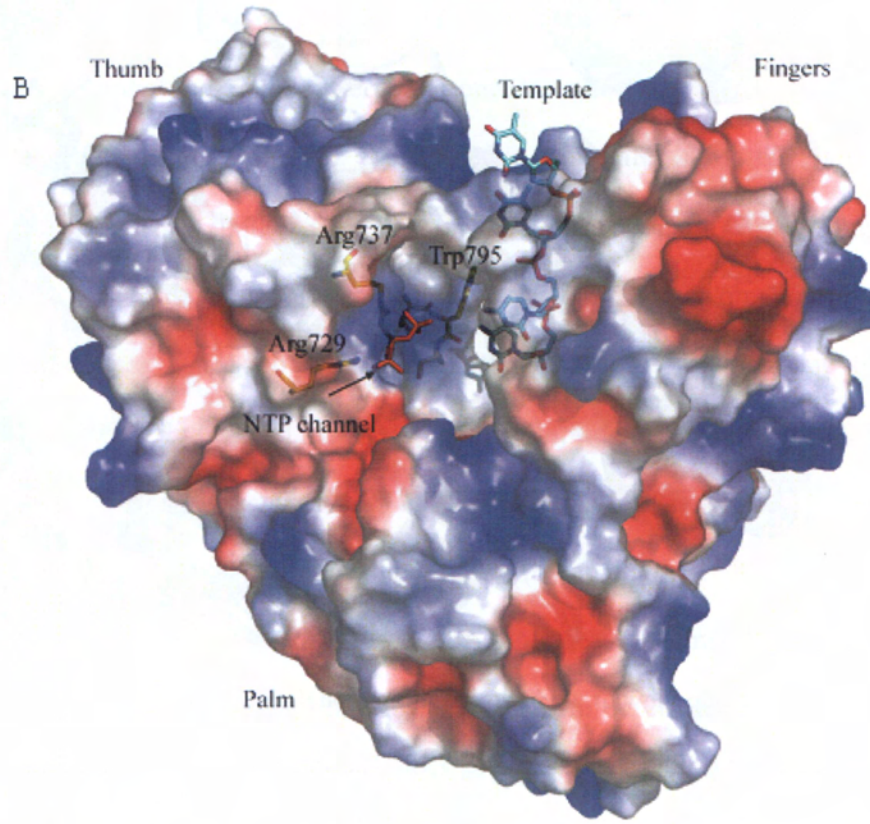


Fig 5.10.Model for the initiation complex of DENV RdRp.

The model is inspired by the Φ6 polymerase quaternary complex and the HCV complex with ribonucleotides. (A) Ribbon representation with the priming base stacking to Trp-795 making Watson-Crick interactions with the template base. (B) Representation of the electrostatic surface of the DENV 3 RdRp viewed from the back. Positive electrostatic potentials are colored blue, and negative potentials are colored red. The view illustrates a ssRNA template and provides a model for the initiation complex. The RNA template traverses the tunnel formed between the thumb and fingers subdomains, entering from the top of the figure. The NTP entry channel is at the interface between the thumb and palm subdomains, indicated by an arrow. The ssRNA template and the tP moiety are shown as sticks (the base and sugar moieties were modeled). Residues discussed in the text that play an important functional role are labeled.

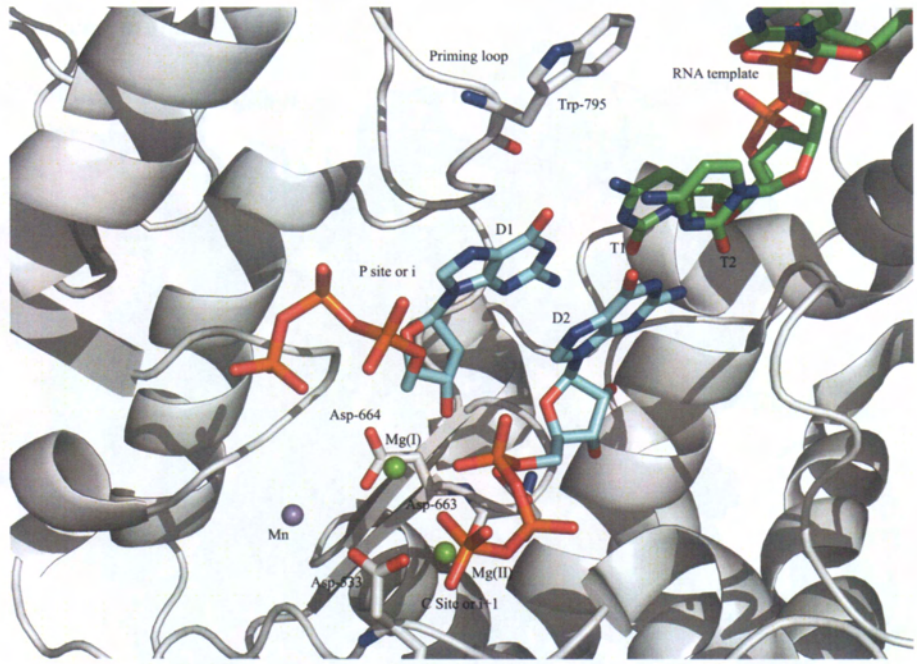


Fig 5. 11. A close-up view of de novo initiation model of DENV 3 RdRp.

The model is a close up view of figure 5.10. The incoming nucleotide D1 is suggested to base stacked on Trp-795. Thus the priming loop conformation could be in closer contact in de novo initiation. Incoming nucleotide D2 stacks on D1. Both incoming nucleotides D1 and D2 base pairs with template T1 and T2, respectively. The incoming nucleotides D1 and D2 are positioned at P site (i) and C site (i+1) respectively. Mg(I) coordinates the α-phosphate and 3'-OH of D1. Whereas Mg(II) interacts with Asp-663, β and γ-phosphate of D2.

The sugar-phosphate backbone of the template is oriented towards the solvent, pointing away from the binding groove of the fingers domain. The positive charge at the surface of the fingers subdomain provides electrostatic interaction with the ssRNA template stabilizing its 3' end at the active site cavity. Together, loops L1, L2 and L3, the linker α21-α22 and the priming loop encircle the catalytic active site and shape the template tunnel. The dimensions of the tunnel we observe are compatible with the presence of only a ssRNA in the active site and not a duplex. Therefore, several loops are likely to adopt a different conformation that

would allow the translocation of a duplex formed by the newly synthesized RNA strand and template. The fact that the tunnel observed in our structure has dimensions that would only permit access to the active site for a ssRNA has implications for the dynamics of the flavivirus polymerase. A motion of helices in the fingers domain and a compression (or closing) of the active site around a bound primer/template has been observed in a number of ternary complexes of DNA polymerases, RNA polymerases and Reverse transcriptases (Doublie et al., 1998; Huang et al., 1998; Li et al., 1998; Lee and Lobigs, 2002). For example, the ternary complex formed between HIV RT and a template:primer:dNTP showed an inward rotation of the fingers subdomain towards the palm (by an angle of ~20°), and the repacking of β-strands β3-β4 (Huang et al., 1998) (or of helices in fingers subdomain for the bacteriophage T7 polymerase), leading to a closed form of the active site in the catalytic complex. In the case of the HCV RdRp template complex (O'Farrell et al., 2003) (PDB code 1NB7), the HCV RdRp apoenzyme structure appears to adopt an energetically favorable closed conformation, and no major conformational changes of the fingers subdomain upon soaking short RNA templates into pre-grown crystal were observed. The high temperature factors in the fingers domains of DENV 3 RdRp suggest that the binding of a short RNA could stabilize this domain and trigger a transition to a more closed form. Compared to the WNV RdRp the DENV 3 RdRp displays an outward rotation with an angle of 8° of the fingers with respect to the palm subdomain. The precise assessment of the importance of conformational changes for the formation of a catalytic complex by flaviviruses RdRps requires further structural studies. It is possible that the polymerase domain expressed separately is more flexible in the absence of the N-terminal MTase region. Alternatively, the flavivirus might represent a separate group (compared to the hepacivirus) where the catalytically competent site is not essentially preformed but instead requires movements of the fingers that would close in the ternary complex. These movements of subdomains are

likely to be accompanied by disorder to order transitions in the segments linking the fingers and thumb domains (e.g. in the linker between helix α4 and α5). Further structural and mutagenesis work is needed to clarify these issues. Interestingly, Gamarnik and collaborators have recently identified a 5' minimal element of 70 bases forming Stem Loop A (SLA), which appears to play an essential role for DENV polymerase RNA binding, leading to template replication (Filomatori et al., 2006). In addition, a brief experimental plan was also proposed for obtaining the ternary-complex structure in DENV RdRp in appendix A4.

5.10 The fidelity model in viral polymerase

In the absence of proofreading activity, the fidelity of viral RdRps is determined by the template-primer-substrate-protein interaction in the active site. A complete kinetic analysis of $3D_{pol}$ (poliovirus) reveals that the nucleotide incorporation can be described by five steps. This provides a kinetic model to understand the polymerase fidelity and role of metal ion (Fig 5.12a) (Castro et al., 2005). These five steps are: the $3D_{pol}$-primer/template (ER_n) binds nucleotide to form a ternary complex (ER_nNTP) and then undergoes a conformational change to form an activated complex that is competent for phosphoryl transfer ($*ER_nNTP$). A ternary product complex ($ER_{n+1}PPi$) is then formed to release PPi after the chemistry reaction occurred via the ($*ER_{n+1}PPi$) as intermediate state. The ER_{n+1} complex are subsequently competent for the next cycle of nucleotide incorporation. The role of residues in DENV RdRp residues that are involved in the fidelity is discussed below, which is based on the proposed model of $3D_{pol}$. Given that the activated ternary complex (Fig 5.12a, step 2) and phosphoryl-transfer step (Fig 5.12a, step 3) that are partially rate-limiting for correct nucleotide incorporation, it is likely that these steps are used to regulate the polymerase fidelity by distinguishing a correct incoming nucleotide from an incorrect nucleotide following Watson-Crick base pairing. Based on the kinetic incorporation of correct and

incorrect nucleotides, Castro et al., 2005 proposed a two step model for nucleotide binding. In the first step, the nucleotide is bound in a "ground state binding" in which the ribose moiety could not bind in an activated orientation. This is because of the Asp of motif A (Asp-238 in $3D_{pol}$ & Asp-538 in DEN RdRp) and Asn of motif B (Asn-297 in $3D_{pol}$ & Asn-609 in DENV RdRp) occlude the ribose sugar binding pocket (see Fig 5.2). In the second step, a conformational change occurs to bring the tp moiety into appropriate position to interact with the conserved active site. This makes the active site ready to accept second metal ion and permit the phosphoryl transfer reaction. Subsequently, the catalysis occurs and releases PPi.

5.10.1 Mn^{2+} as mutagenic metal ion

In BVDV and HCV, the RNA synthesis occurred *in vitro* only when both Mn^{2+} and Mg^{2+} were provided exogenously. As discussed earlier, Mn^{2+} reduces the K_m for the initiation by GTP and increases de novo initiation product. Thus it was proposed that the substitution of Mn^{2+} for Mg^{2+} decreases fidelity. In $3D_{pol}$ kinetic study, the use of Mn^{2+} as divalent ion cofactor could diminish the rate of phosphoryl transfer for incorrect nucleotide relative to correct nucleotide, compared to Mg^{2+} (Fig 5.12b) (Arnold and Cameron, 2004; Arnold et al., 2004). The interaction between the $3D_{pol}$-Mn^{2+} increases the stability of the activated ternary complex, compared to 3D-Mg^{2+} (Fig 5.12b, step 2). Thus the authors proposed that Mn^{2+} increases the stability of the activated ternary complex, regardless of the correct or incorrect base or sugar configuration, by binding more tightly to the β and γ of the tp moiety than the Mg^{2+} and makes the orientation of the tp moiety independent of interactions with residues in the ribose-binding pocket (Arnold et al., 2004). Therefore, the inability to couple the nature of the bound nucleotide to the efficiency of phosphoryl transfer is the primary reason for the observed loss of $3D_{pol}$ fidelity in the presence of Mn^{2+} (Castro et al., 2005).

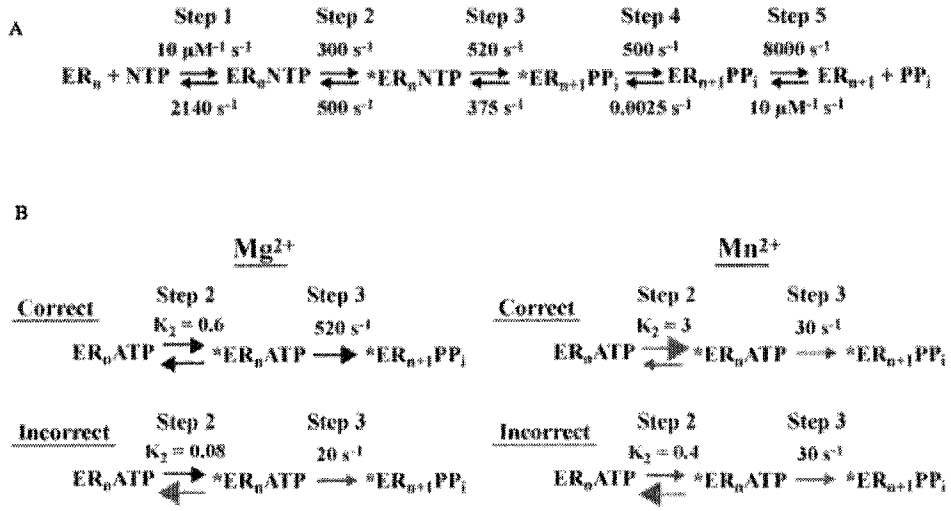


Fig 5.12.Schematic kinetic for $3D_{pol}$.

(A) Complete kinetic mechanism for 3Dpol-catalyzed nucleotide incorporation. (B) Comparison between the conformational change step and the phosphoryl transfer step for $3D_{pol}$ catalyzed correct and incorrect nucleotide incorporation in the presence of Mg^{2+} and Mn^{2+}. Adapted from Catro et al., 2005.

5.10.2. Key residue in fidelity: Superimposition of DENV RdRp and $3D_{pol}$

Interrogation of the $3D_{pol}$ nucleotide binding pocket revealed 6 residues that are in the vicinity of the nucleotide substrate, including Asp-233, Asp-238, Ser-288, Thr-293, Asn-297 and Asp-328 (Fig 5.13) (Gohara et al., 2004; Gohara et al., 2000). To identify the residues that play fidelity role in DENV RdRp replication, we superimposed these protein structures (DENV RdRp and $3D_{pol}$). These residues were identified and correspond to Asp-533, Asp-538, Ser-600, Thr-605, Asn-609 and Asp-663 in DENV 3 RdRp (Fig 5.13). In addition, the HIV-RT-primer/template-substrate complex was used by the authors to develop a model for the $3D_{pol}$ complex (Gohara et al., 2000). With kinetics and structural model, they found Asp-238 (Asp-538 in DENV) in motif A is the key determinant of RdRp fidelity. Residue Asp-

238 and Asn-297 of $3D_{pol}$ (corresponding to Asp-538 and Asn-609 in DENV RdRp) was observed to have no contact with the ribose of the nucleotide in the apoenzyme (Fig 5.13). This model is referred as ground binding state. A conformational change occurs to orientate the tp moiety ready for phosphoryl transfer. The Asp-238 and Asn-297 are re-positioned to accommodate the ribose and make interaction with the nucleotide. The ribose is held by the interaction between the 3'-OH and the backbone of Asp-238 and by interaction between the 2'-OH and Asn-297 (Fig 5.13). This orientation of the complex allows the alignment of the second divalent metal ion that permits the release of PPi. The disruption of interaction between Asn-297 and 2'-OH of ribose may also alter the interaction between the 3'-OH and the β-phosphate of the nucleotide substrate, resulting in tp moiety movement. Consequently, the efficiency of nucleotide incorporation also reduces (Castro et al., 2005). This model is consistent with a higher free energy being required for incorrect nucleotide incorporation (dAMP), compared to the correct nucleotide (AMP) (Gohara et al., 2004). These results and model are also in good agreement with the FMDV RdRp elongation complex with different nucleotide that was discussed earlier (see Fig 5.2). In the case of mispair, the orientation of tp moiety is altered by the non polar mispair. Therefore, stabilization of the tp moiety-metal complex in principal is contributed by the conserved motif that provides interaction network, the fundamental of nucleotide incorporation for polymerase. Based on this model, these authors further proposed an universal and adapted mechanism of nucleotide incorporation and phosphoryl transfer (Fig 5.14).

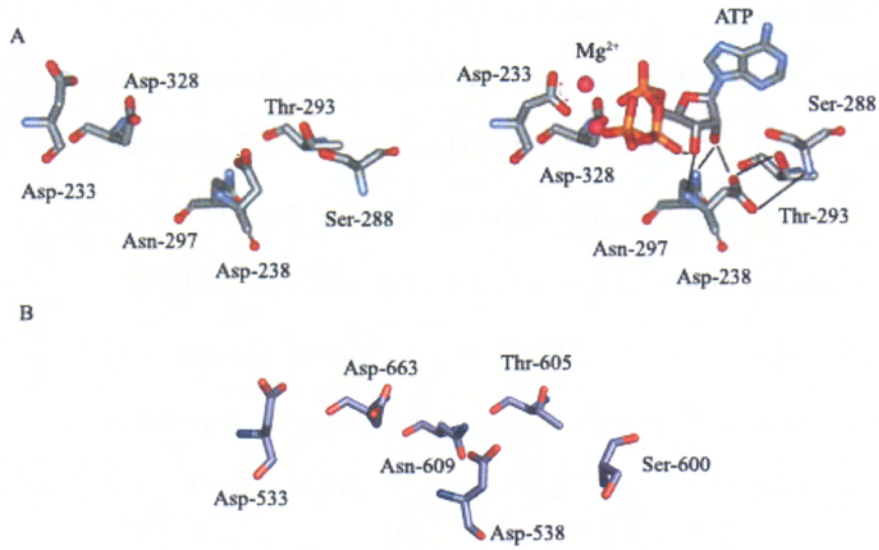


Fig 5.13.Nucleotide binding pocket.

(A) Nucleotide binding pocket of $3D_{pol}$. (Left) residues located in the NTP binding pocket, Asp-233 and Asp-238 are from motif A while Ser-288, Thr-293 and Asn-297 are from motif B, Asp-328 is from motif C. (Right) $3D_{pol}$ interaction model with bound nucleotide. In this model, the side chains for Asp-233 and Asp-293 have been orientated for interaction. (B) Nucleotide binding pocket of DENV 3 RdRP that identified from superimposition with $3D_{pol}$. The residue Asp-533 and Asp-538 are from motif A, Ser-600, Thr-605 and Asn-609 are from motif B, Asp-663 is from motif C.

In the canonical palm that contains the nucleotide binding pocket, motif A (Asp-233 and Asp-238 of $3D_{pol}$ and Asp-663 and Asp-538 of DENV 3 RdRp) serves as universal mechanism for metals and tp moiety catalysis. This mechanism is structurally conserved to DENV RdRp as well. The disruption on residues in motif A could results in the perturbation on position of both sugar and tp moiety and reduce the efficiency of phosphoryl transfer. Whereas motif B (Asn-297 in $3D_{pol}$ or Asn-609 in DENV) (Fig 5.13) serves as adapted site for respective polymerase. The residue Asn-297 enable enzyme to differentiate ribose from

the 2'-deoxyribose. This site is not conserved across polymerase but generally is conserved in RdRp. Together the kinetic and structural model demonstrated the role of residue Asp-238 and Asn-297 and suggests that Asp-538 and Asn-609 in DENV RdRp also play a similar role. In HIV-RT, this adapted site is Phe-160 that has van der Walls interaction with the 2'-OH. The Phe-160 alters the Tyr-115 orientation (motif A) in the presence of the 2'OH, likewise motif B of DNA polymerase (T7, Taq and KF) also cause movement of motif A via the residue located in the sugar binding pocket. The discovery of residues that modulate RdRp fidelity has valuable information in drug design. Small molecules can be designed to alter the enzyme fidelity, either by increasing the incorporation fidelity that make the virus susceptible to selection pressures that evades the virus or reduce the incorporation fidelity that leads virus to catastrophe due replication error.

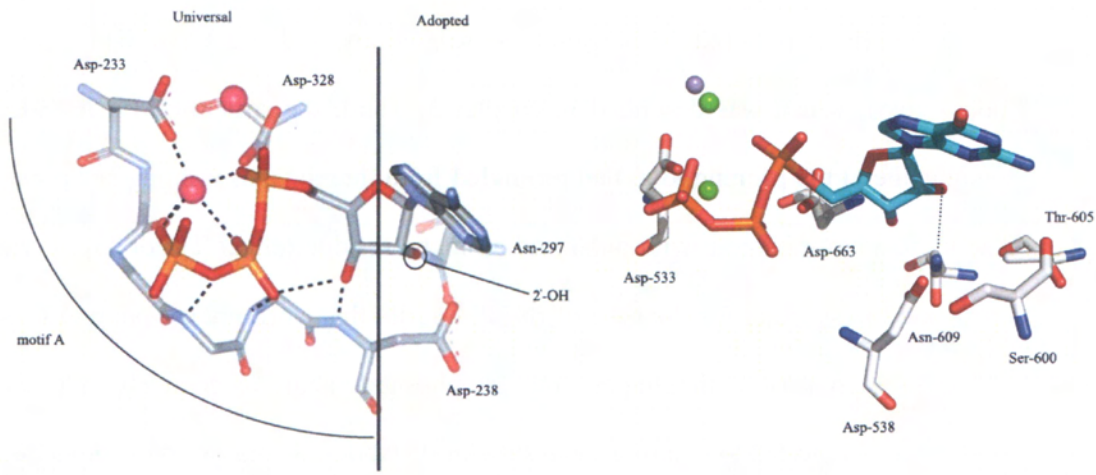


Fig 5.14.Conserved mechanism for binding a correct nucleotide.

The conserved mechanism for $3D_{pol}$ model (Left) and DENV 3 RdRp (Right). This mechanism was proposed by Cameron et al., 2004. The nucleotide-binding pocket of all polymerases with a canonical "palm" based active site is highly conserved. The site is divided into 2 parts: (I) universal interaction mediated by conserved motif A that coordinate metals and tp moiety for

catalysis. (II) Adapted interaction mediated by conserved motif B that dictates ribo or 2' deoxyribonucleotides is utilized. The Asn-609 of DENV 3 RdRp is conserved in flaviviral RdRp. The Mg^{2+} is represented in ball in magneta and green respectively while Mn^{2+} is in purple color.

5.11 Conclusion

In this work, we elucidated the crystal structure of DENV 3 RdRp which shed light on the role and function of the RdRp catalytic domain. First, the architecture of the protein assumes the canonical right-hand conformation consisting the fingers, palm and thumb, which is characteristic of known polymerase structure. Second, the structure clearly revealed that the NLS region is not a separate region. The NLS region is part of the polymerase structure and is believed not only to mediate nuclear localization of NS5 but also to facilitate the open-close conformation during the formation of productive polymerase-template-primer complex, similar to those of fingertips of HCV RdRp. This is also partly explained from the fact that protein construct required NLS region for soluble and stable protein expression (273-900 amino acid), which was described in chapter 3. Third, the structure also revealed a rNTP binding site at the priming loop that protruded from thumb. This priming site plays a critical role in de novo initiation, with polar and aromatic residue (either Tyr or Trp in viral RdRp) providing the stacking for the base of rNTP. Fourth, this work unexpectedly discovered two Zn binding pockets in the fingers and thumb subdomains respectively. The Zn binding pockets are suggested to contribute the structural stability. Fifth, two Mn^{2+} ions binding sites were found in DENV 3 RdRp. The first Mn^{2+}(I) is located at the catalytic pocket, which is in good agreement with metal A position that proposed in "two metal ions" mechanism. This Mn^{2+} is coordinated by Asp-663 of motif C. The second metal binding site (both Mn^{2+} & Mg^{2+}) is located at a site which is ~6 Å from the catalytic pocket (similar to Φ6 RdRp). This allosteric metal binding site is unique to viral RdRp (RHDV and Φ6 RdRp). Nevertheless, no

metal ion was found in the position B. Similar observation was found in the case of FMDV active elongation complex. The reason for the absence of metal ion at position B in DENV 3 RdRp is unclear. However, the presence of substrate/template in the crystals will allow us to accurately assess the effect of Mn^{2+} on the finger subdomains motion. A model for DENV ssRNA synthesis was also proposed. This model revealed that the template tunnel only allows the active site to access a ssRNA. Lastly, this work has also discussed the structural basis of fidelity of viral RdRp. The fidelity model led us to identify key residues and understand how these residues play critical roles in modulating the DENV RdRp fidelity. Residue Asn-609 and Asp-538 are identified as key residues that determine the fidelity of RdRp, based on the kinetic model on $3D_{pol}$ and FMDV RdRp complex structure. The residue Asp-538 of DENV 3 RdRp is proposed to enable enzyme to differentiate ribose from the 2'-deoxyribose. This valuable information is useful in drug design to design small molecules that can alter the enzyme fidelity.

CHAPTER 6

SUMMARY AND FUTURE DIRECTION

6.0 Summary

6.0.1 Lessons from protein expression

The soluble NS5 related proteins expression was discussed. Initial efforts were focused on crystallizing the DENV 2 FLNS5 (TSVO1 strain) that was expressed using the baculovirus system. After extensive crystallization campaign, no crystal growth condition was encountered. Efforts to express truncated proteins that do not contain the NLS also led to no success as these proteins were insoluble and aggregated into inclusion bodies. The 3D structure of the DENV RdRp that was obtained in this study clearly demonstrated that the NLS is not a separate domain as previously thought and is integral to the RdRp domain and its absence will affect expression and solubility. This study highlights one of the pitfalls of predicting domain boundaries based on domain boundaries arising from functional studies with truncated constructs. With the aim to increase the likelihood of obtaining soluble, stable and good yield of proteins for functional and structural studies, protein refolding, limited proteolysis and homologous protein screening were employed. Following limited proteolysis studies with DENV 2 FLNS5, crystallization trials with DENV 3 RdRp (273-900; which also includes the NLS) found a suitable crystal growth condition. Its important to note that the protein which yielded the crystal was identified from the homologous protein screening, compared to DENV 2 TSVO1. Initially the FLNS5 crystallization was the focus of this work and much effort were put into screening on the WNV wild type NS5 or mutant, co-crystallization of WNV FLNS5-FLNS3 and DENV 2 FLNS5-Fab C3 fragments. In summary, none of these attempts were successful and this underscores the challenges in crystallizing the

FLNS5 protein, which presumably is due to the high mobility between the MTase and RdRp domains.

It is obvious from the preceding discussions that the lessons learned from the protein production are valuable. Often, protein engineering is applied in areas such as chemical, pharmaceutical and agricultural industries to design protein with new and improved properties. Such approaches to facilitate preparation of protein production and high quality crystals have been widely applied in protein crystallography and reviewed by Derewenda (Derewenda, 2004). Two general strategies can be adapted. The first strategy is rational design, which requires extensive knowledge of the protein structure (DiTursi et al., 2006; Hayes et al., 2002). This information may not be available for most projects at the initial stage. Several rational modification approaches that are based on protein structure (secondary or tertiary) being already available. Rational site directed mutagenesis, is an approach to modify surface residues or surface patches that mediate (i) crystal contacts, (ii) crystal resolutions or (iii) solubility/ thermo-stability. Other approaches such as removal of flexible regions including termini and interior loops can also be applied under rational design where flexible regions can be predicted either by computational methods or biochemical studies such as protease accessibility. The second strategy is directed evolution, which use multiple round of random mutagenesis (can be error prone PCR) or DNA shuffling to generate protein variants (Stemmer, 1994a; Stemmer, 1994b). In this study, we concluded that the combination of (1) protein domain identification with biochemical method, limited proteolysis studies and (2) homologous protein screening to generate protein variants is the optimal strategy. This combined strategy is effective in achieving (1) best protein constructs for optimal soluble protein production and (2) proteins that can readily be crystallized. In short, despite the identification of the protein domain (residue 273-900), only DENV 1 & 3

RdRp were found to be stable in term of protein solubility and purification yield, compared to DENV 2 & 4 RdRp. To underscore this, only DENV 3 RdRp produced crystals after crystallization screening using both DENV 1 & 3 RdRp. This suggests that homologous protein screening is a highly efficient approach and makes use of protein sequences that are already represented in the natural population of microorganisms. The application of this strategy is also validated by the successes in our laboratory for other viral proteins that have yielded high quality crystals in our lab (MTase, DENV FLNS3, NS3 helicase). Thus, the combination of limited proteolysis and homologous protein screening strategy can be considered as an effective approach to characterize proteins that are derived from other viruses in the future. In the case of full length protein, neither rational design that is based on rational surface residue entropy prediction to generate variants (WNV) nor the use of homologous proteins were successful in crystallizing the protein. Again, this failure can be due to many factors. Nevertheless, homologous protein screening helped us to obtain best yield of NS5 related full length proteins. In our attempts to generate stable DENV 3 RdRp, we have succeeded to identify a surface polar residue that affect protein stability based on rational protein engineering with the crystal structure. We mutated residue 846 (appendix A2, table 2) and this allows us to have a better thermo-stable protein to further characterize protein using biophysical methods such as NMR and ITC. To conclude, the strategies we have adopted can be valuable lesson for structural studies of new viral protein.

6.0.2 Aspects of the enzymatic efficiency of RdRp

The lower enzymatic activity of DENV RdRp compared to the FLNS5 could be attributed to the template specificity. DENV RdRp might be less sensitive to the homopolymeric RNA (poly C) than the FLNS5. As discussed earlier, DENV RdRp has been shown to bind 5' UTR

and 3' UTR of the dengue genome. The use of natural template such as the SLA region in the 5' UTR or 3' UTR may be used to optimize the sensitivity of the enzymatic activity.

The 3D structure determined in this work suggests that the open and flexible conformation of DENV RdRp compared to RdRp structure of WNV (Malet et al., 2007) may impact on enzymatic activity. It is possible that in the FL NS5 a more closed structure may be adopted. Currently, we have no way to estimate how open is the conformation of DENV RdRp compared to RdRp of FLNS5 since the FLNS5 crystal structure is not available. The advent of several biophysical methods permits the exploration of protein dynamics, and the hypothesized conformational difference in RdRp likely can be examined in detail. Small angle X-ray scattering, which deduces overall globular structure of the protein in solution is one such method. It is likely that of the comparison of the solution structure of RdRp catalytic domain vs. FLNS5 can provide the answer. This technique is discussed more below (see future outlook). Single molecule fluorescence resonance energy transfer (sm FRET), is a second feasible method that allows the characterization of individual molecule beyond ensemble-averaged properties. This approach uses donor and acceptor dyes that are attached to two sites of a molecule can be used to measure the distance between the two dyes. The measurement will help us to estimate the approximate distance, for example between the fingers and thumb domain during catalysis for the RdRp vs. FLNS5. In addition, the measurement of a change in fluorescence from the two dyes can serve as an indicator of a change in the conformation of the molecule. Nevertheless, the challenge is the difficulty in identifying suitable residues in the molecule for labeling. The set up of this approach is also useful for probing DENV RdRp fidelity mechanisms. Some successful studies have been applied on DNA polymerase (Bailey et al., 2001; Luo et al., 2007). One study observed a non rate limiting conformational change before phosphodiester bond formation which serves as a

kinetic checkpoint to discriminate against the incorporation of non complementary NTPs (Luo et al., 2007). Similarly, this approach can be used to validate the proposed DENV RdRp fidelity model (chapter 5). In addition, this sm FRET can be used as complementary tool to validate hits from HTS and provide more detailed mechanistic information on action of the inhibitor. This is because sm FRET can measure the detailed time trajectories of molecular conformations (Luo et al., 2007). Traditional biochemical enzymatic assays use higher concentration of DENV NS5 protein than what is present in a dengue infected cell. In addition, traditional enzymatic characterization that is represented by an ensemble population of proteins, may have possibly lost some of the important information such as intramolecular motion that is represented by a sub population of proteins. Therefore, sm FRET approach that characterizes single molecule (much lower protein concentration is used) which likely represents the copy number of proteins in the "real environment" (infected cells), may provides some insightful idea to explain why certain strains of DENV RdRp are more virulent than others. A separate study has shown a correlation of viral genome sequence and dengue pathogenesis (Leitmeyer et al., 1999). Full genome sequences of 11 DENV were analyzed and structural differences were seen consistently between those associated with dengue fever and those with potential to cause DHF, which reside in regions that include NS5. It is also known that studies of wild and attenuated dengue viruses have suggested that genetic differences among strains of the four serotypes can be associated with attenuation, virulence and epidemic potential(Hanley et al., 2002). These ambiguities will be answered in the future when more dynamic information as well as structural information become available for the different serotypes.

6.0.3 Insights from structural studies

Overall perspective

The crystal structure of DENV 3 RdRp which shed light on the role and function of the RdRp catalytic domain was described. The architecture of the DENV 3 RdRp protein assumes the canonical right-hand conformation consisting of fingers, palm and thumb, characteristic of known polymerase structures, but with the NLS region playing an important role in the formation of the observed structure. Interestingly, the NLS domain signatures described previously (Brooks et al., 2002) are distributed between the fingers and thumb subdomains. Three helices (α2- α4) are incorporated within the thumb and the fingers (α 5- α7) subdomains respectively and the connection between them is realized through the segment linking helices α4 and α5. The truncations within the NLS region are thus likely to destabilize the protein structure, an observation which explains *a posteriori* the poor expression/solubility of constructs lacking part of this region. The structure of DENV 3 RdRp also clearly shows that the role of the NLS region is most probably to facilitate the open-close conformation during the polymerization. This suggestion is based on structural studies of a NNI-HCV RdRp complex which depicted how NNI binds at the surface of the thumb domain. In the apoenzyme, this site is occupied by an α-helix of the fingertips (similar to NLS in our case) that connects the finger and thumb domains. The binding of NNI blocks the connection between the fingers and thumb domain and the fingertips can no longer mediate the concerted polymerization steps (Wang et al., 2003). Interestingly, a segment of 20 amino-acids that is strictly conserved in all flaviviruses also lies between amino acid residues 320-368 of NLS region (Fig 6.1). This region of NS5, which is involved in binding importin β, is also thought to interact with NS3 protease-helicase (Fig 6.1). Structural elucidation on how NS3 interacts with NLS in NS5 while this region facilitates open-close conformation in concert during polymerization, will offer great insight both for replication events and drug

discovery efforts. The thumb domain is the most diverse feature among all the known viral RdRps. The thumb domain of Flaviviridae RdRps contains 2 times more residues than that of Picornaviridae RdRps. As described earlier, DENV 3 RdRp has a priming loop that protrudes from the thumb into the active site that reduces the template tunnel. This may allow only ssRNA access to the active site during initiation. A small positively charged tunnel on the back side of the thumb is postulated to modulate nucleotide diffusion. Palm domain of DENV 3 RdRp that contains an anti-parallel β sheet appears the most structurally conserved among all known polymerases. This reflects the catalytic site architecture is preserved during evolution. Four conserved motif that are responsible for NTP binding and catalysis have been identified in the palm domain, including the GDD catalytic active site.

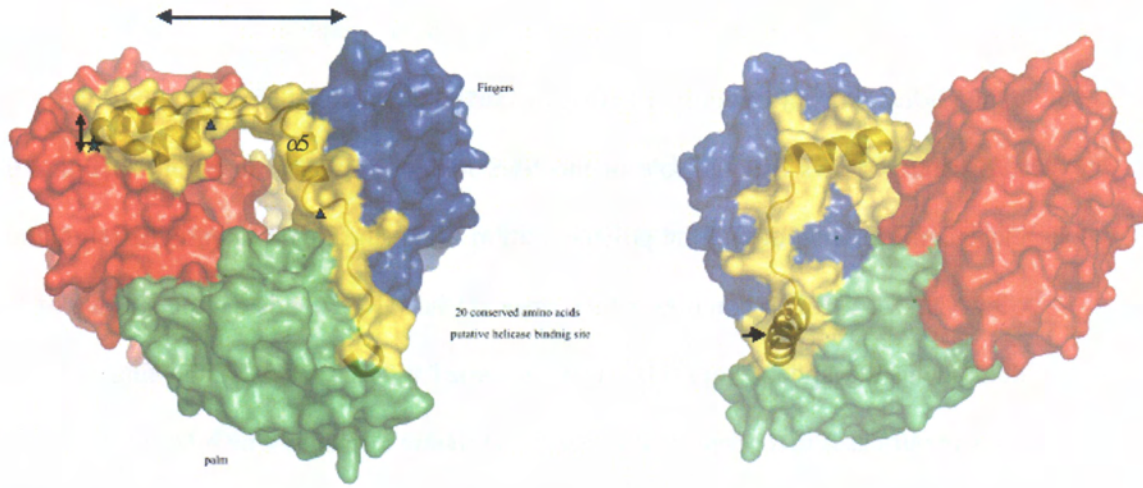


Fig 6. 1. Surface representation for position of putative binding sites on DENV 3 RdRp.

(Left): Back view of the DENV 3 RdRp. Fingers, thumb and palm domains are represented by blue, red and green color, respectively. The bNLS region is represented by yellow color. The region of 20 conserved amino acids that was proposed to bind NS3 helicase or importin-b is represented by triangle and near α5. The bNLS region also forms the fingertips which is similar to that of HCV RdRp. The fingers-thumb motion for open-close conformation may involves α4- α5

of the fingertips. The arrows represent the potential motion for the concerted formation of elongation complex. It will be interesting to know how NS3 interacts with NLS region during polymerization. The star icon represents the similar location where a NNI of HCV RdRp was shown to block the motion of fingertips duing polymerization. (Right): Front view of the DENV 3 RdRp. The a/b NLS region is shown in yellow color. Residues corresponding to DENV 2 basic cluster that necessary for nuclear import, are indicated by an arrow.

Kinetic mechanism from a structural perspective

Previous structural studies of several viral RdRps from (+) and ds RNA viruses indicate a highly conserved general feature of RdRp architecture, despite the fact that these structures arise from a diverse range of viruses. The arrangement of fingers, palm and thumb subdomains, as well as the motifs A-F are universally conserved. Therefore, the enzymatic mechanism of polymerization is thought to be highly conserved as seen in the structural conservation in RdRps. As discussed in chapter 5, kinetic analysis of $3D_{pol}$ underlying the nucleotidyl incorporation which follows a five steps reaction cycle: (1) the binding of an NTP that is complementary to the base in the template to form an open complex; (2) a conformational change of the enzyme-template-substrate to the closed complex; (3) nucleotidyl transfer and translocation; (4) a second conformational change and; (5) finally the release of PPi. Experimentally trapping these different states of kinetic steps are challenging for structural studies. Nevertheless, there are several breakthrough cases, include (1) Φ6 RdRp-RNA-NTP complexes that appear to represent step 1 of the enzymatic mechanism (the initiation complex), (2) FMDV RdRp-RNA-NTP complexes represent the open complex at step 2 or 5 (elongation complex) and (3) Norwalk virus RdRp-RNA-NTP complexes represent the step 3 of closed complex (Zamyatkin et al., 2008). Based on these respective complexes, different kinetic steps of enzymatic mechanism were modeled into DENV 3

RdRp (Fig 6.2). In addition, the initiation of replication from single stranded RNA template to the exit of duplex RNA was also modeled into DENV 3 RdRp (Fig 6.3 & 6.4).

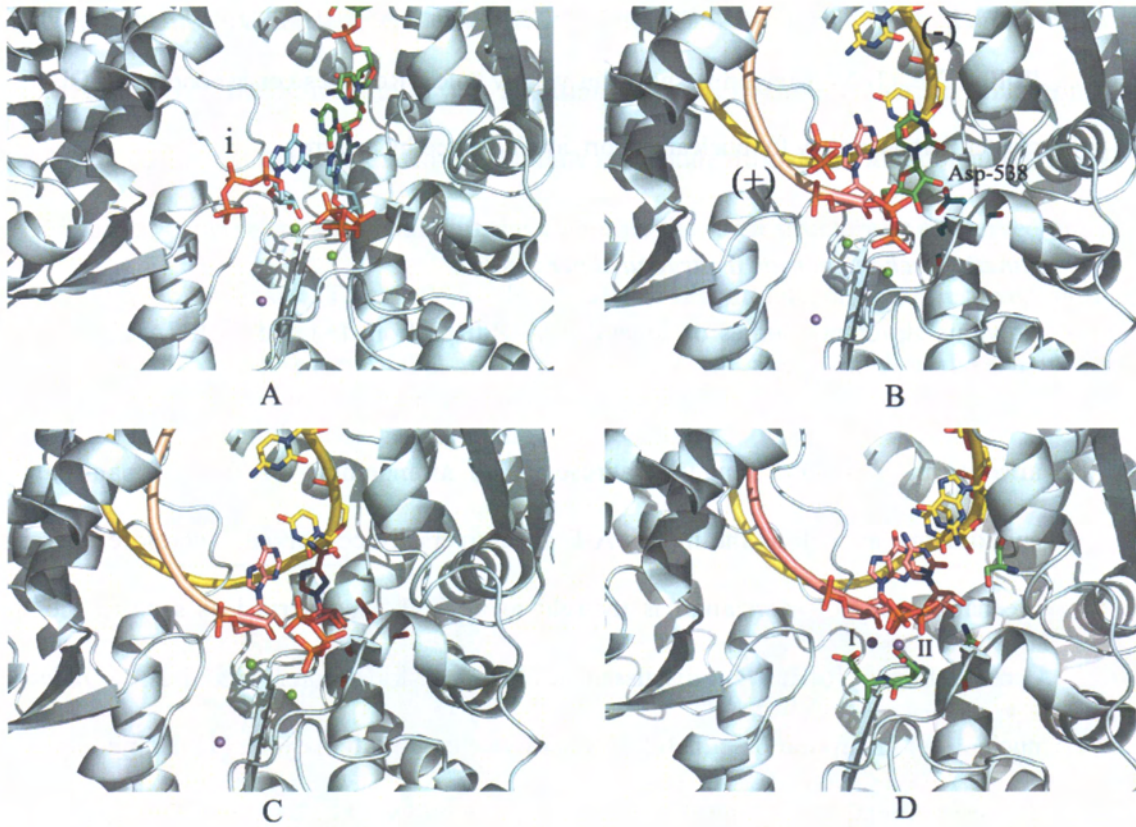


Fig 6. 2. Close-up view of kinetic mechanism steps for DENV 3 RdRp.

Different kinetic steps enzymatic mechanism for polymerase were modeled into DENV 3 RdRp. All figures are close up view at the active site from the back of RdRp. The (-) RNA template is used for the synthesis of (+) RNA in this model. The first nucleotide for the 3' of newly synthesized RNA is shown in stick. (A) Initiation replication of DENV 3 RdRp (step 1), which is proposed based Φ6 RdRp. Details please see text. Substrate NTP i (left) is base stacked with residue Trp-795 that is located at priming loop while substrate NTP at catalytic site i+1 is regulated by metal ions and catalytic residues. Template is shown in stick and green color. (B) and (C) elongation complex of DENV 3 RdRp (step 2 or 5), which is proposed based on FMDV

RdRp. The ribose moiety of UTP in Fig B is in ground state binding (not bind in an optimal orientation) due to Asp-538 and Asn-609 occlude the ribose binding pocket. Residues Asp-538 and Asn-609 are reoriented to stabilize the position of ribose by direct interactions formed with the incoming nucleotide in Fig C. Complex in Fig B also represents the release of PPi of step 5. Template is shown in yellow color and new strand of RNA is in orange color. (D) Closed complex of DENV 3 RdRp, which is proposed based on NV RdRp. Compared to Fig D, complexes in Fig B and C are in open conformation due to the NTP bound too far from the primer terminus for the nucleotidyl transfer reaction and a catalytic metal ion (II) is missing from the active site. In addition, the incoming nucleotide in Fig D is base stacked with the first nucleotide at the 3' of the newly synthesized RNA (orange color).

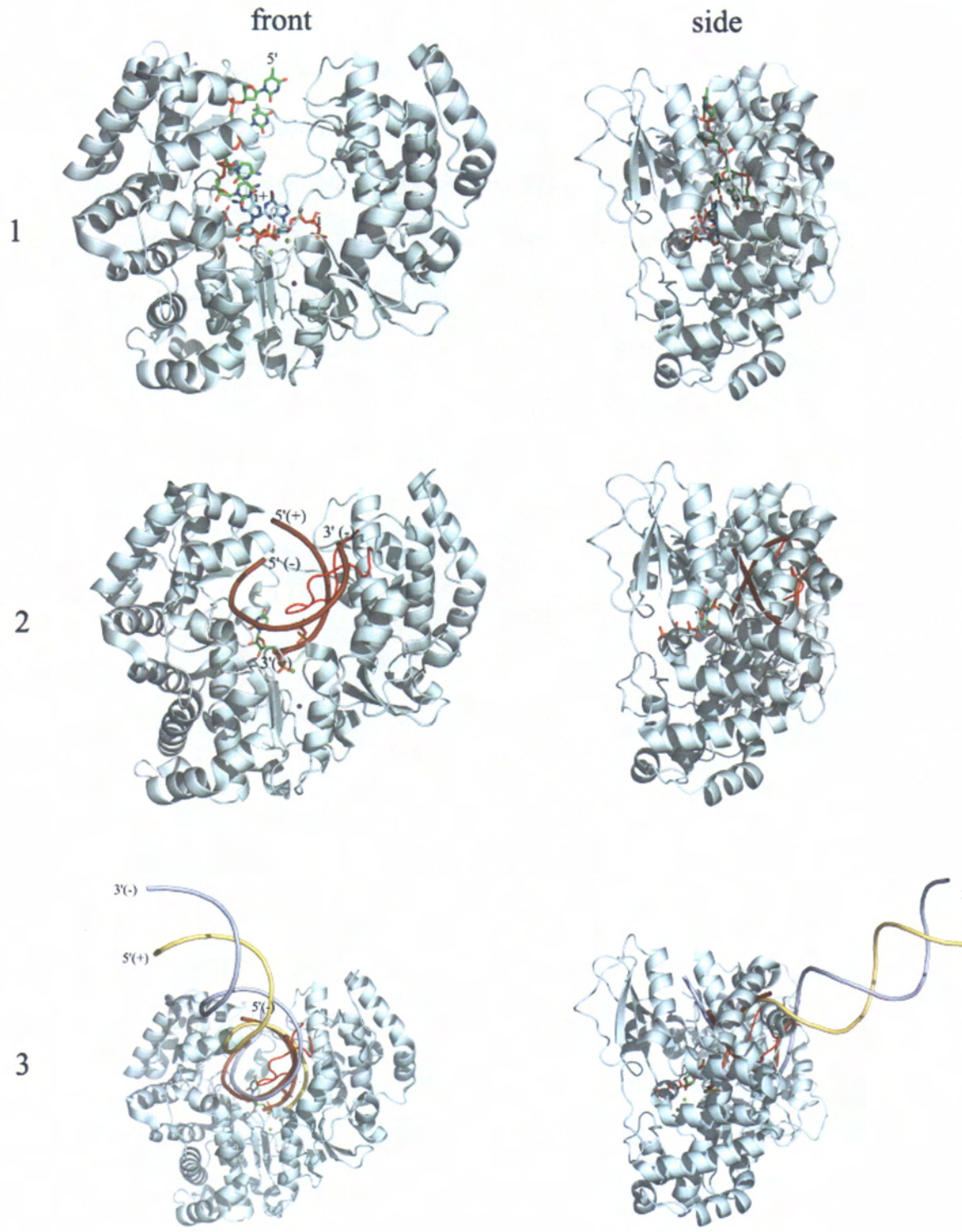


Fig 6. 3. A model for the exit of duplex RNA(dsRNA) from DENV 3 RdRp.

The dsRNA exit model for DENV 3 RdRp was built based on Φ6 RdRp, FMDV RdRp and HIV-RT, respectively that represent step 1-3. In this case, the (-) ssRNA is used as template for the synthesis of (+) RNA. (1) The (-) ssRNA template (green) enters the RdRp from the top and the 3' of the template accesses the catalytic site. (2) The dsRNA exits from the front tunnel of DENV RdRp. Several loops such as L1-L3 and priming loop encircle the catalytic site and shape the RNA template tunnel. The template tunnel is only compatible with ssRNA in the active site. In this model, priming loop (red color) is modeled to adopt different conformation that allow the exit of dsRNA. (3) The conformation for the exit of longer dsRNA is shown, which is built based on HIV-RT. The (-) RNA template and newly synthesized (+) RNA are shown in light purple and yellow color, respectively. Interestingly, the dsRNA conformation that both resides in the template tunnel for FMDV (brown)and HIV-RT is similar. This enables the model to propose how likely is the conformation of dsRNA that has exited and out of the RNA template tunnel of RdRp.

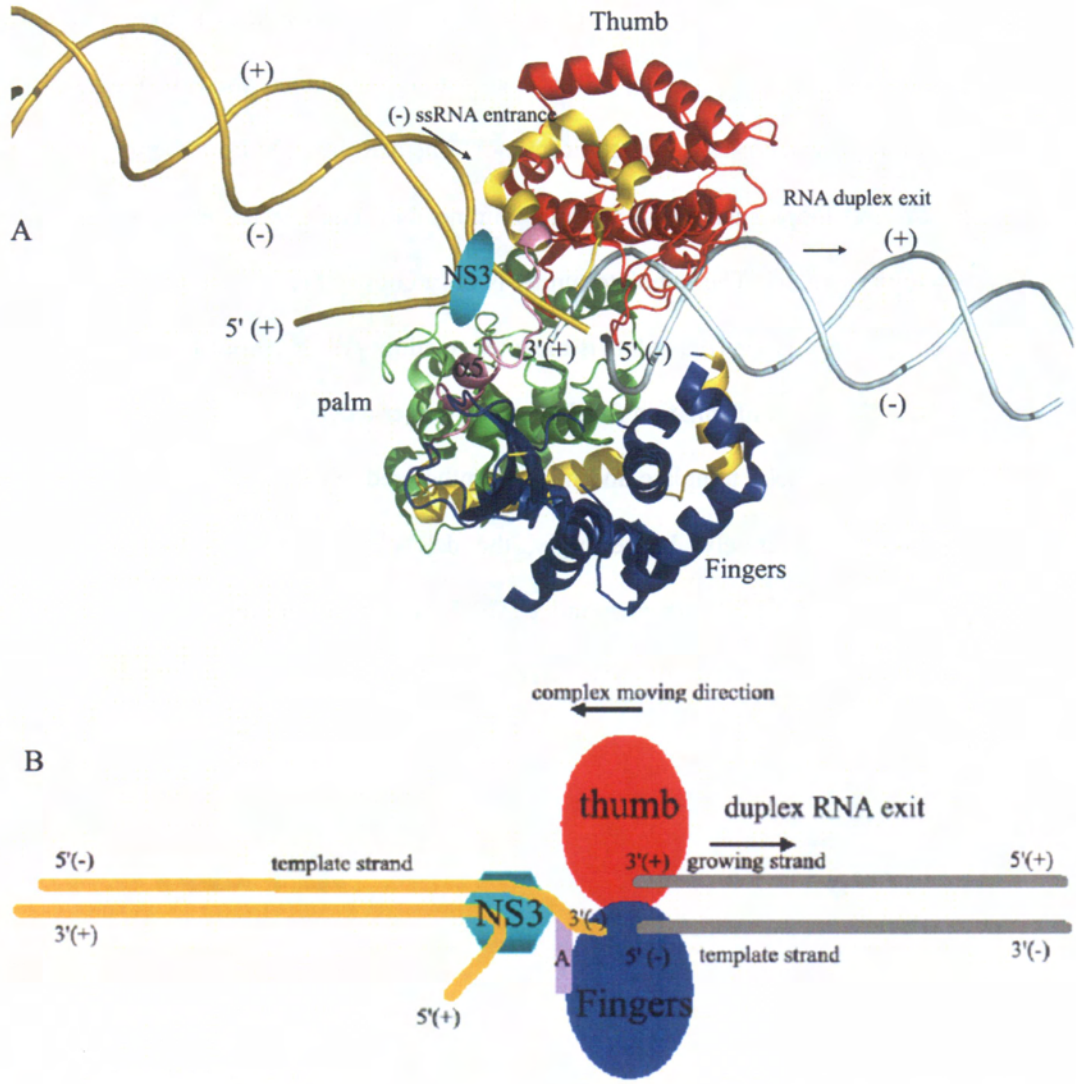


Fig 6. 4. Replication model for DENV 3 RdRp.

The replication model is viewed on the top. This model is proposed based on (I) 20 conserved amino acids of bNLS or near $\alpha 5$ of RdRp was suggested to interact with NS3 helicase during RNA replication (Brooks et al., 2002; Forwood et al., 1999), (II) the exit of dsRNA model in Fig 6.3. The synthesized dsRNA (grey) exits from the front of RdRp (right side), while the NS3 that interact with A region (purple color, the 20 conserved amino acids) to unwind the dsRNA in order to allow the 3' of the (-) ssRNA accesses catalytic site.

Step 1 of kinetic mechanism: de novo initiation view from Φ6 RdRp complex

The Φ6 RdRp-RNA-NTP initiation complexes (ssRNA) show that the first nucleotide (D1) forms a Watson-Crick base pairing with the base (T1) of template at 3' end. The D1 is base stacked on Tyr-630 (P site or i). The second nucleotide of Φ6 (D2) stacks on D1 and base pairs with the T2 of template. The D2 is also positioned and coordinated by two Mg^{2+} ions at the catalysis site (C site or i+1). The template base changes orientation in response to the RNA oligonucleotides binding template. The T1 is rotated ~ 180°, T2 is stabilized by base stacking with T3. These changes results in the change of sugar conformation and side chains of interacting polymerase residues, leading to the extra hydrogen network that further stabilizes the RdRp-RNA-substrate complex, compared to DNA-polymerase complex.

Based on Φ6 RdRp, a de novo initiation complex model for a DENV was also proposed (Fig 6.2A). The position of RNA template, GTP and the catalytic ions were inferred from the Φ6 initiation complex. The sugar-phosphate backbone of the template is oriented towards the solvent, pointing away from the binding groove of the fingers domain of DENV RdRp. The positive charge at the surface of the fingers subdomain provides electrostatic interaction with the ssRNA template stabilizing its 3' end at the active site cavity. Structural elucidation shown that loops L1, L2 and L3, the linker of α21- α22 and the priming loop encircle the catalytic site and shape the RNA template tunnel. The model revealed that the template tunnel is only compatible with ssRNA in the active site. Therefore, the structure is likely to adopt a different conformation that would allow the translocation of a duplex formed by the newly synthesized RNA strand and template. DENV 3 RdRp structure also revealed a rNTP (3'dGTP) binding site at the priming loop that protruded from the thumb, 7 Å away from the catalytic site (C site or i+1). This priming site is similar to that of P site (i) of Φ6. The orientation of the triphosphate tp moiety (3'dGTP) is consistent to allow its base to forms a

stacking interaction with the Trp-795 (similar to that of GTP at P site or i, near Tyr-630 of Φ6). This priming site plays a critical role in de novo initiation. The tP moiety of 3'dGTP is coordinated by residues that are conserved across the positive strand RNA viruses that are known to use de novo initiation. Mutational analysis further supports the essential role of the conserved residues on de novo synthesis (Lai et al., 1999). A high GTP concentration was required for the formation of de novo initiation complex, regardless of the sequence of the first 6 nucleotides at 3'end template. It was hypothesized that the GTP play a key role (1) GTP is the initiating nucleotide for (-) strand RNA synthesis and (2) GTP has putative allosteric binding site to allow the enzyme to adopt a closed conformation for de novo initiation.

Step 2 or 5 of kinetic mechanism: elongation and fidelity view from FMDV RdRp complexes

FMDV RdRp-RNA-NTP elongative complexes (duplex RNA) show the RNA elongation process and fidelity of RNA replication (Ferrer-Orta et al., 2004). The structure demonstrates that the translocation of the RNA product and positioning of the new incoming substrate UTP close to the active site. The UTP is in the ground state binding in which the ribose moiety could not bind in an optimal orientation because Asp-245 (motif A) and Asn-307 (motif B) occlude the ribose binding pocket. These key residues are structurally conserved both in DEN and $3D_{pol}$ (corresponding to Asp-538 and Asn-609 in DENV, Asp-238 and Asn-297 in $3D_{pol}$). In contrast, RdRp-RNA-RTP complex shows a conformational change for Asp-245 and Asn-307. These two residues are reoriented to stabilize the position of ribose by direct interactions formed with the 3' and 2'-OH groups of the incoming nucleotide (Fig 6.2B and 6.2C). The active site is in a correction position for the phosphoryl transfer reaction. The 2'-OH of RTP forms a double hydrogen bond with the side chains of Asn-307 and Asp-245. Furthermore, the 3'-OH forms a hydrogen bond with the catalytic Asp-338 and the N- atom of the main

chain of Asp-245. These structures further corroborated the model proposed by Cameron et al., 2004 that suggesting that these residues may play a role in the control of nucleotide incorporation and fidelity. These two complexes could represent the second step of the kinetic mechanism.

In the second step, the key residues of RdRp-RNA-RTP complex are re-orientated to the active position to form interaction with in the incoming nucleotide. However, the complex appears to be open complex with the RTP bound too far from the primer terminus for the nucleotidyl transfer reaction to occur and one of the two catalytic metal ions is missing from the active site. In addition, FMDV-RNA-ATP complex displays the formation of a new base pair and translocation of the RNA product (represents step 5) (Fig 6.2B).

To understand the structural basis for the fidelity of DENV 3 RdRp, a model that is based on the kinetic studies of $3D_{pol}$ and FMDV elongation complexes was proposed (Fig 6.2B and 6.2C). Superimposition of DENV 3 RdRp and $3D_{pol}$ has helped us to identify residue Asp-533 and Asp-538 from motif A, Ser-600, Thr-605 and Asn-609 from motif B, and Asp-633 from motif C are nucleotide binding pocket in DENV 3 RdRp. Further analysis shows that residue Asn-609 and Asp-538 are the key residues that modulate the fidelity of viral RdRp (similar to that of [1] Asn-297 and Asp-238 of $3D_{pol}$ and [2] Asn-307 and Asp-245 of FMDV). It is likely Asp-538 and Asp-609 of DENV 3 RdRp is re-positioned to accommodate the ribose to make interaction with the nucleotide. The ribose is held by the interaction between the (1) 3'-OH and the backbone of Asp-538 and (2) 2'-OH and the Asn-609. Asn-609 is also likely the key residue that enables the enzyme to differentiate ribose from the 2'-deoxyribose. Thus, Asn-609 and Asp-538 are identified as the key residues that modulate the fidelity of DENV 3 RdRp. It is our hopes that the proposed residues that modulate RdRp fidelity can be

used to design small molecules. The small molecules can either increase the incorporation fidelity that make virus susceptible to selection pressures which evades the virus or reduce the incorporation fidelity that leads to viral replication error.

Comparison of four different FMDV complex structures in the active site reveals common contacts mediated by residues of motifs C and E of the palm domain that stabilize the position of the 3' end of the primer strand in an orientation for RNA elongation. Motif A, B of the palm and motif F in fingers were also found to help the positioning of the acceptor base of the template strand and participate in the recognition and binding of the incoming rNTP. Interestingly, motif F position in DENV RdRp-RNA template binding is different from FMDV. This could probably be due to evolutionary pressure to distinguish specific protein-RNA template interaction. The position of Ser-298 and Thr-303 that are involved in template and incoming nucleotides interaction are critical in virus replication. This was shown by mutagenesis (Ferrer-Orta et al., 2007). Interestingly, these residues are also structurally conserved in DENV RdRp (Ser-600 and Thr-605) and suggest the conserved role of these residues.

Step 3 of kinetic mechanism: the closed complex

Norwalk virus (NV) RdRp-RNA-NTP complexes showed that RNA duplex and NTP bound in the active site cleft. The binding of RNA also causes the central helix of the thumb domain to rotate by 22°, thus forming a binding groove for the primer strand. This rotation appears to be coupled with the C-terminal tail moving away from the active site cleft. These conformational changes are reminiscent of changes seen in a number of ternary complexes of polymerases (Doublie et al., 1998; Huang et al., 1998; Li et al., 1998). For example, the ternary complex formed between HIV-RT and a template:primer:dNTP showed an inward

rotation of the fingers subdomain towards the palm (by an angle of ~20°), and the repacking of β-strands β3-β4 (Huang et al., 1998), leading to a closed form of the active site in the catalytic complex. This observation in combination with the modes of binding for catalytic metal ions and NTP, further suggest that the NV RdRp-RNA-NTP complex is a closed complex, which is trapped before the nucleotidyl transfer. The NTP bases form hydrogen bonds with the complementary GTP base of the template and stack against the 3' terminal base of the primer, as well as the Arg-182 of motif F. Although this Arg is highly conserved in RdRps, the position of motif F in the NV RdRp structure is different from that of DENV 3 RdRp. This situation is similar when compared to FMDV RdRp. Again, the 2'-OH and 3'-OH of NTP make a similar hydrogen bonding pattern (Asn-309 and Asp-247) with the conserved residues as seen in FMDV and $3D_{pol}$. However, the FMDV-RdRp-RNA-NTP complex may not in the closed conformation due to the fact that the RTP is bound too far from the primer terminus for the nucleotidyl transfer reaction to occur and also one of the two catalytic metal ions is missing from the active site. In NV-RdRp complexes, two Mn^{2+} ions were found to coordinate three highly conserved Asp residues and the NTP tp moiety for the catalysis through two metal ion mechanism (Fig 6.2D).

6.0.3 Two metal ions mechanism

Two metal ions mechanism was proposed by Steitz (Steitz, 1998) for polymerase- catalyzed nucleotidyl transfer reactions. Divalent ions were shown to coordinate NTP, catalyze the formation of phosphodiester bond and transiently stabilize the by-product PPi. Metal ion B contacts the γ and β-phosphate of the incoming nucleotide and Asp residue from motif A and water molecules, is likely to stabilize the oxyanion that forms in the penta-covalent transition state geometry and to facilitate the release of pyrophosphate. During catalysis, metal B orients the tP moiety in the active site and may contribute to charge neutralization. Once the

nucleotide is in place, metal ion A is bound and coordinated by the 3'-OH of the primer, α-phosphate of NTP and Asp residues from motifs A and C. Metal ion A is likely to lower the pK_a of the primer 3'-OH group of the primer and form the $3'O^-$ to promote nucleophilic attack on the α-phosphate at physiological pH.

The NV-RdRp complexes provide structural evidence on how the two metal ion mechanism works in viral RdRp. The complexes reveal that metal ion A octahedrally coordinates the 3'-OH nucleophile of the primer, the NTP α-phosphate, the carboxylate groups of Asp residues from motif A and C as well as water molecule. By donating hydrogen bonds to the negatively charged primer phosphodiester and NTP α-phosphate group, the water molecule could serves as general base catalyst. This allowing a proton to leave from the 3'-OH primer and then followed by nucleophilic attack on the α-phosphate. Nevertheless, the observed water molecule role in NV-RdRp complexes is somehow different from a separate study on poliovirus RdRp ($3D_{pol}$). The study revealed that 2 protons transfer occurs in the nucleotidyl transfer transition state (Castro et al., 2007) and provides evidence that (1) a Lys-359 of $3D_{pol}$ to serves as a general acid for protonation of the PPi and (2) Asp-328 and Asp-329 serve as general base to play a role in primer 3'-OH deprotonation in the polymerase-catalyzed nucleotidyl-transfer reactions. However, the authors do not rule out the use of water molecule as general base for deprotonation of the 3'-OH. During the nucleotidyl transfer reaction, the proton from the primer 3'-OH is removed and donated to the leaving PPi (Castro et al., 2007). Lys-689 of DENV 3 RdRp is located to a position similar to the general acid Lys-359 of $3D_{pol}$. The proposed general base is catalytic conserved Asp residues (Asp-663 and Asp-664 in DENV RdRp).

6.0.4 Metal ion binding at active site

Several studies show that the use of Mn^{2+} in HCV RdRp enzymatic reaction could increases the de novo initiation product and reduces K_m for the initiation nucleotide GTP (Ranjith-Kumar et al., 2002). A separate study by Castro et al., 2005 also proposed that nucleotide binding incorporation to RdRp is driven by the metal liganded triphosphate moiety of nucleotide at the active site. A conformational change occurs to bring the metal bound tp moiety into appropriate position to interact with the conserved aspartyl groups of the polymerase. Kinetic studies on $3D_{pol}$ suggest that Mn^{2+} could increases the stability of the activated and closed ternary complex by binding tightly to the β and γ of the tP moiety, regardless of the nature (correct or incorrect) of the base or sugar conformation. The orientation of the tP moiety is independent of the interactions with residues in the ribose binding pocket, thus allowing Mn^{2+} to diminish the rate of phosphoryl transfer for incorrect nucleotide relative to correct nucleotide, compared to Mg^{2+}.

In this study, two Mn^{2+} ions binding site were also found in DENV 3 RdRp. The first Mn^{2+}(I) is located at the catalytic pocket (coordinated by Asp-663 and Asp-664) while the second metal binding site (Mn^{2+} & Mg^{2+}) is located at a site which is ~6 Å from the catalytic pocket, a site similar to that of Φ6 RdRp. This allosteric metal binding site is unique to viral RdRp. Asp-533 of DENV 3 RdRp was found to interact with metal ion (Mn^{2+} or Mg^{2-}) which located at the allosteric pocket via water molecule. In WNV RdRp, Asp-536 and Asp-669 coordinate a Ca^{2+} ion at the similar allosteric site (corresponding to Asp-533 and Asp-664 in DENV RdRp) (Malet et al., 2007). Mutation of this residue in BVDV (Asp-533 in DENV RdRp) abolishes the elongative synthesis and significantly reduces the de novo synthesis (Lai et al., 1999), indicating the conserved role of this residue in DENV 3 RdRp for RNA replication. On other hand, structural analysis on the catalytic Asp residues (motif A and C)

in Φ6 RdRp and FMDV complex also shows that these residues are not in optimal geometry (>3 Å) for regulating the metal ions, compared to the tight metal ions-Asp catalytic residues interaction (< 2 Å) seen in HIV-RT ternary complex. Further experiments are needed to assess the effect of Mn^{2+} and Mg^{2+} binding at the catalytic site and polymerization.

6.0.5 Zinc binding sites

Unexpectedly two Zn binding pockets in the fingers and thumb subdomains were discovered. The Zn binding pockets is suspected to play a role in the structural stability given its location in the structure. This is a novel observation as the other Flaviviridae RdRp do not possess this feature. Residues that are involved in the Zn binding pocket (Zn2) which is located at the anti- parallel β-sheet in thumb domain are completely conserved in four serotypes of DENV and YFV. Interestingly, this site is not conserved in WNV RdRp. This Zn2 is postulated to regulate conformational switches within the thumb domain. In contrast, residues that coordinate Zn binding pocket (Zn1) and located in the fingers subdomains are conserved also in WNV RdRp. Further experiments are needed to address the exact role of these Zn binding sites.

6.1 Outlook

There are many uncertain aspects and unanswered questions concerning the DENV polymerase. Chiefly, the DENV RdRp model in this study is derived from current crystal form($C222_1$) that has several residues missing which may be involved important function. They are residues: (1) 311-316, between β1-α2 a fingertip that regulates the entrance of RNA template which access the active site, (2) 406-418, between α7- α8, a loop that appears to regulate the exit of RNA template and (3) 451-469, between α9- α10, a loop that partially contains the motif F sequence. In addition, the current crystal condition requires air

dehydration and crystal growth at 4 ºC in a month. These crystal handling (air dehydration and growth condition) procedures are troublesome for X-ray crystallographic screening with compound library. Therefore, surface mutagenesis on residues involved in crystal contact can be employed. This strategy can serve: (1) to identify new conditions in which crystals grow in a shorter time and without the need for further crystal treatment. The condition also should be easy and friendly for compound library soaking studies, (2) to identify a crystal condition that show the missing residues in current model. This is also equally important for compound screening, if these missing residues have important interaction with the potential compounds. In addition, 2D NMR screening is recommended to be incorporated in the X-ray crystallographic screening to improve the hit rate.

The crystal structure of DENV 3 RdRp was elaborated in chapter 5. In the future, we need to understand the enzyme-RNA-template-rNTP (substrate) ternary complex from a structural standpoint. The ternary complex could provide the structural basis for understanding the polymerase activity, either during the de novo initiation or elongation process. The complex also should provide insight into the molecular basis of fidelity for viral polymerase. For example, the FMDV polymerase elongation complexes that were snapshot at different successive elongation process could define critical residues involved in: (1) recognition and positioning of the incoming nucleotide, (2) positioning of the acceptor base of the template strand and (3) the positioning of 3'OH of the primer nucleotide during RNA replication. In addition, the ternary complex crystal can serve as an important tool for drug resistance study, especially for the nucleoside analog. In the case of HIV-RT, the ternary crystal structure in complex with AZT (nucleoside analog) that trapped the AZT at pre-translocation (at nucleotide binding site) and the post-translocation (at priming site) respectively, offer molecular details relevant to the mechanism of ATP-dependent HIV-RT resistance to AZT

(Sarafianos et al., 2002). In a different prospect, structural basis of transition from initiation to elongation, translocation of RNA template, rNTP substrate and exiting of the duplex strand from the Flaviviridae polymerase remain to be unraveled. Structural elucidation of these various stages of replication could promote better understanding on the role and function of polymerase in viral replication and for drug design. Similar different stages of replication have been solved in T7 RNA polymerase (Steitz, 2006).

In the full length NS5 context, currently there is no structural data available. As discussed in chapter 3, we suspected that the RdRp catalytic domain alone adopts more open and flexible conformation. Also that the FLNS5 was found to be more active than the RdRp. Thus how the role and function of RdRp catalytic domain is modulated by the presence of MTase and NLS region in the context of full length NS5 in replication event remains to be answered. Based on data from the reverse genetic studies as a guide and by using the docking approach, Malet et al., 2007 proposed a WNV FLNS5 model (see Fig 6.5). They found residue 512 Leu-->Val was compensated when K46/R47A/E49A of MTase was mutated in the replicon system (Malet et al., 2007). The resulting model revealed that residues Lys-46, Arg-47 and Glu-49 loop of MTase fits well into a large groove located between the thumb and the fingers domains and adjacent to the Leu-512 loop. However, this model was not based on direct empirical data. Therefore, the model could deviate from the actual conformation state of FLNS5.

Due the difficulty of crystallizing the FLNS5, we propose to use small angle X-ray scattering (SAXS) to solve the FL model. Briefly, SAXS is a solution technique that can determine biological sample shape by *ab initio* and domain structure. The detailed modeling of biological macromolecular complexes is then proceeded by using rigid-body refinement

(refer Svergun and Koch, 2002 for detail). The resolution of this approach is up to 1.5 nm and can be applied on samples ranging from 5 kDa to 100 MDa in solution. Therefore this approach offers a new direction for obtaining the FL model. We hope this approach can lead us to answer questions and ambiguity that require explanation for the FLNS5 structure. The availability of specific antibodies against DENV 2 FLNS5 also can be used to obtain the FL structure via this approach. Another advantage in using this approach is that only a relatively a small quantity of sample is required. Nevertheless, the biological sample is required in monodispersed form for SAXS determination. Recently, the Lescar laboratory in collaboration with professor Grűber (NTU) has also determined the SAXS structure of FLNS3, which is in good agreement crystal structure (Luo et al., 2008) thus providing some proof for the reliability of this approach.

The structure obtained in this work is a very important starting point for detailed understanding of the DENV RdRp. The use of multitude of biophysical techniques proposed here as well as the availability of small molecule inhibitors that will result from the drug discovery efforts will be important tools for detailed probing of the structure and mechanism of action.

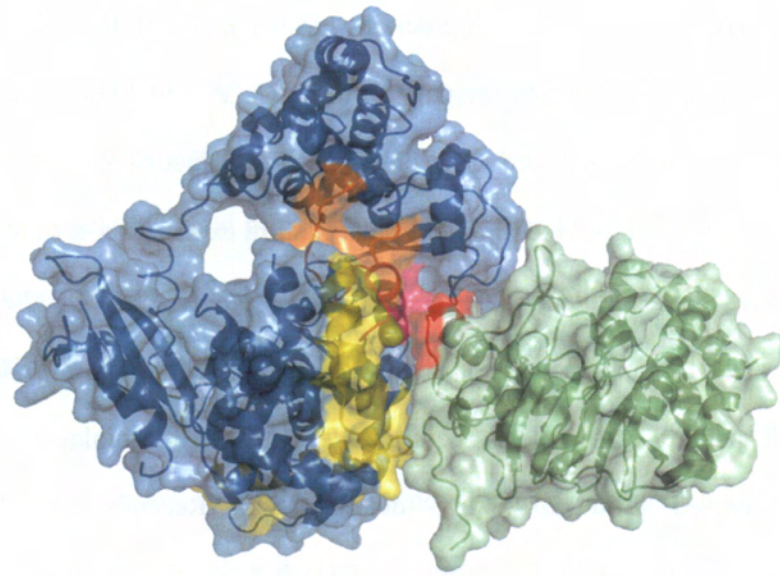


Fig 6. 5. Full length model of DENV NS5.

The model is generated based on Malet. H et al., 2007. The RdRp is in blue color, with the NLS is in yellow color. The MTase is in green color. The K46/K47/E49 of MTase that predicted interacts with L512 of RdRp are shown in red and pink color respectively.

Reference List

Ackermann,M. and Padmanabhan,R. (2001). De novo synthesis of RNA by the dengue virus RNA-dependent RNA polymerase exhibits temperature dependence at the initiation but not elongation phase. Journal of Biological Chemistry *276*, 39926-39937.

Adachi,T., Ago,H., Habuka,N., Okuda,K., Komatsu,M., Ikeda,S., and Yatsunami,K. (2002). The essential role of C-terminal residues in regulating the activity of hepatitis C virus RNA-dependent RNA polymerase. Biochimica et Biophysica Acta-Proteins and Proteomics *1601*, 38-48.

Adamson,C.S. and Freed,E.O. (2008). Recent progress in antiretrovirals - lessons from resistance. Drug Discovery Today *13*, 424-432.

Ago,H., Adachi,T., Yoshida,A., Yamamoto,M., Habuka,N., Yatsunami,K., and Miyano,M. (1999). Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. Structure *7*, 1417-1426.

Ahlquist,P., Noueiry,A.O., Lee,W.M., Kushner,D.B., and Dye,B.T. (2003). Host factors in positive-strand RNA virus genome replication. Journal of Virology *77*, 8181-8186.

Alcon,S., Talarmin,A., Debruyne,M., Falconar,A., Deubel,V., and Flamand,M. (2002). Enzyme-linked immunosorbent assay specific to dengue virus type 1 nonstructural protein NS1 reveals circulation of the antigen in the blood during the experiencing primary acute phase of disease in patients or secondary infections. Journal of Clinical Microbiology *40*, 376-381.

Allison,S.L., Schalich,J., Stiasny,K., Mandl,C.W., Kunz,C., and Heinz,F.X. (1995). Oligomeric rearrangement of tick-borne encephalitis virus envelope proteins induced by an acidic pH. J. Virol. *69*, 695-700.

Alvarez,D.E., Lodeiro,M.F., Luduena,S.J., Pietrasanta,L.I., and Gamarnik,A.V. (2005). Long-range RNA-RNA interactions circularize the dengue virus genome. J. Virol. *79*, 6631-6643.

Amberg,S.M., Nestorowicz,A., Mccourt,D.W., and Rice,C.M. (1994). Ns2B-3 Proteinase-Mediated Processing in the Yellow-Fever Virus Structural Region - In-Vitro and In-Vivo Studies. Journal of Virology *68*, 3794-3802.

Appleby,T.C., Luecke,H., Shim,J.H., Wu,J.Z., Cheney,I.W., Zhong,W., Vogeley,L., Hong,Z., and Yao,N. (2005). Crystal structure of complete rhinovirus RNA polymerase suggests front loading of protein primer. J. Virol. *79*, 277-288.

Arakawa,T. and Timasheff,S.N. (1985). The Stabilization of Proteins by Osmolytes. Biophysical Journal *47*, 411-414.

Area,E., Martin-Benito,J., Gastaminza,P., Torreira,E., Valpuesta,J.M., Carrascosa,J.L., and Ortin,J. (2004). 3D structure of the influenza virus polymerase complex: Localization of

subunit domains. Proceedings of the National Academy of Sciences of the United States of America *101*, 308-313.

Arias,C.F., Preugschat,F., and Strauss,J.H. (1993). Dengue-2 Virus Ns2B and Ns3 Form A Stable Complex That Can Cleave Ns3 Within the Helicase Domain. Virology *193*, 888-899.

Arnold,J.J. and Cameron,C.E. (2004). Poliovirus RNA-dependent RNA polymerase (3D(pol)): Pre-steady-state kinetic analysis of ribonucleotide incorporation in the presence of Mg2+. Biochemistry *43*, 5126-5137.

Arnold,J.J., Gohara,D.W., and Cameron,C.E. (2004). Poliovirus RNA-dependent RNA polymerase (3D(pol)): Pre-steady-state kinetic analysis of ribonucleotide incorporation in the presence of Mn2+. Biochemistry *43*, 5138-5148.

Au,K., Berrow,N.S., Blagova,E., Boucher,I.W., Boyle,M.P., Brannigan,J.A., Carter,L.G., Dierks,T., Folkers,G., Grenha,R., Harlos,K., Kaptein,R., Kalliomaa,A.K., Levdikov,V.M., Meier,C., Milioti,N., Moroz,O., Muller,A., Owens,R.J., Rzechorzek,N., Sainsbury,S., Stuart,D.I., Walter,T.S., Waterman,D.G., Wilkinson,A.J., Wilson,K.S., Zaccai,N., Esnouf,R.M., and Fogg,M.J. (2006). Application of high-throughput technologies to a structural proteomics-type analysis of Bacillus anthracis. Acta Crystallographica Section D-Biological Crystallography *62*, 1267-1275.

Auwerx,J., Stevens,M., Van Rompay,A.R., Bird,L.E., Ren,J.S., De Clerq,E., Oberg,B., Stammers,D.K., Karlsson,A., and Balzarini,J. (2004). The phenylmethylthiazolylthiourea nonnucleoside reverse transcriptase (RT) inhibitor MSK-076 selects for a resistance mutation in the active site of human immunodeficiency virus type 2 RT. Journal of Virology *78*, 7427-7437.

Bailey,M.F., Thompson,E.H.Z., and Millar,D.P. (2001). Probing DNA polymerase fidelity mechanisms using time-resolved fluorescence anisotropy. Methods *25*, 62-77.

Bailey,S. (1994). The Ccp4 Suite - Programs for Protein Crystallography. Acta Crystallographica Section D-Biological Crystallography *50*, 760-763.

Balint,A., Baule,C., Kecskemeti,S., Kiss,I., and Belak,S. (2005). Cytopathogenicity markers in the genome of Hungarian cytopathic isolates of bovine viral diarrhoea virus. Acta Veterinaria Hungarica *53*, 125-136.

Ball,L.A. (1995). Requirements for the Self-Directed Replication of Flock House Virus Rna-1 (Vol 69, Pg 721, 1995). Journal of Virology *69*, 2722.

Baneyx,F. and Mujacic,M. (2004). Recombinant protein folding and misfolding in Escherichia coli. Nature Biotechnology *22*, 1399-1408.

Barre-Sinoussi,F., Chermann,J.C., Rey,F., Nugeyre,M.T., Chamaret,S., Gruest,J., Dauguet,C., Axler-Blin,C., Vezinet-Brun,F., Rouzioux,C., Rozenbaum,W., and Montagnier,L. (2004). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Revista de Investigacion Clinica *56*, 126-129.

Bartelma,G. and Padmanabhan,R. (2002). Expression, purification, and characterization of the RNA 5 '-triphosphatase activity of dengue virus type 2 nonstructural protein 3. Virology *299*, 122-132.

Bartholomeusz,A.I. and Wright,P.J. (1993). Synthesis of Dengue Virus-Rna Invitro - Initiation and the Involvement of Proteins Ns3 and Ns5. Archives of Virology *128*, 111-121.

Baudin,F., Bach,C., Cusack,S., and Ruigrok,R.W.H. (1994). Structure of Influenza-Virus Rnp .1. Influenza-Virus Nucleoprotein Melts Secondary Structure in Panhandle Rna and Exposes the Bases to the Solvent. Embo Journal *13*, 3158-3165.

Bazan,J.F. and Fletterick,R.J. (1989). Detection of A Trypsin-Like Serine Protease Domain in Flaviviruses and Pestiviruses. Virology *171*, 637-639.

Bedard,K.M. and Semler,B.L. (2004). Regulation of picornavirus gene expression. Microbes and Infection *6*, 702-713.

Benzaghou,I., Bougie,I., and Bisaillon,M. (2004). Effect of metal ion binding on the structural stability of the hepatitis C virus RNA polymerase. Journal of Biological Chemistry *279*, 49755-49761.

Biswal,B.K., Wang,M., Cherney,M.M., Chan,L., Yannopoulos,C.G., Bilimoria,D., Bedard,J., and James,M.N. (2006). Non-nucleoside inhibitors binding to hepatitis C virus NS5B polymerase reveal a novel mechanism of inhibition. J. Mol. Biol. *361*, 33-45.

Blight,K.J. (2007). Allelic variation in the hepatitis C virus NS4B protein dramatically influences RNA replication. Journal of Virology *81*, 5724-5736.

Bosch,J., Robien,M.A., Mehlin,C., Boni,E., Riechers,A., Buckner,F.S., Van Voorhis,W.C., Myler,P.J., Worthey,E.A., DeTitta,G., Luft,J.R., Lauricella,A., Gulde,S., Anderson,L.A., Kalyuzhniy,O., Neely,H.M., Ross,J., Earnest,T.N., Soltis,M., Schoenfeld,L., Zucker,F., Merritt,E.A., Fan,E., Verlinde,C.L.M.J., and Hol,W.G.J. (2006). Using fragment cocktail crystallography to assist inhibitor design of Trypanosoma brucei nucleoside 2-deoxyribosyltransferase. Journal of Medicinal Chemistry *49*, 5939-5946.

Boutayeb,A. (2006). The double burden of communicable and non-communicable diseases in developing countries. Transactions of the Royal Society of Tropical Medicine and Hygiene *100*, 191-199.

Boyer,P.L., Sarafianos,S.G., Arnold,E., and Hughes,S.H. (2000). Analysis of mutations at positions 115 and 116 in the dNTP binding site of HIV-1 reverse transcriptase. Proceedings of the National Academy of Sciences of the United States of America *97*, 3056-3061.

Boyer,P.L., Sarafianos,S.G., Arnold,E., and Hughes,S.H. (2001). Selective excision of AZTMP by drug-resistant human immunodeficiency virus reverse transcriptase. Journal of Virology *75*, 4832-4842.

Bressanelli,S., Tomei,L., Rey,F.A., and De Francesco,R. (2002). Structural analysis of the hepatitis C virus RNA polymerase in complex with ribonucleotides. J. Virol. *76*, 3482-3492.

Brinton,M.A. (2002). The molecular biology of West Nile virus: A new invader of the Western hemisphere. Annual Review of Microbiology *56*, 371-402.

Brinton,M.A. and Dispoto,J.H. (1988). Sequence and Secondary Structure-Analysis of the 5'-Terminal Region of Flavivirus Genome Rna. Virology *162*, 290-299.

Brinton,M.A., Fernandez,A.V., and Dispoto,J.H. (1986). The 3'-Nucleotides of Flavivirus Genomic Rna Form A Conserved Secondary Structure. Virology *153*, 113-121.

Brooks,A.J., Johansson,M., John,A.V., Xu,Y.B., Jans,D.A., and Vasudevan,S.G. (2002). The interdomain region of dengue NS5 protein that binds to the viral helicase NS3 contains independently functional importin beta 1 and importin alpha/beta-recognized nuclear localization signals. Journal of Biological Chemistry *277*, 36399-36407.

Bruenn,J.A. (2003). A structural and primary sequence comparison of the viral RNA-dependent RNA polymerases. Nucleic Acids Research *31*, 1821-1829.

Butcher,S.J., Grimes,J.M., Makeyev,E.V., Bamford,D.H., and Stuart,D.L. (2001). A mechanism for initiating RNA-dependent RNA polymerization. Nature *410*, 235-240.

Cahour,A., Pletnev,A., Vazeillefalcoz,M., Rosen,L., and Lai,C.J. (1995). Growth-Restricted Dengue Virus Mutants Containing Deletions in the 5' Noncoding Region of the Rna Genome. Virology *207*, 68-76.

Calisher,C.H., Monath,T.P., Karabatsos,N., and Trent,D.W. (1981). Arbovirus Subtyping - Applications to Epidemiologic Studies, Availability of Reagents, and Testing Services. American Journal of Epidemiology *114*, 619-631.

Castro,C., Arnold,J.J., and Cameron,C.E. (2005). Incorporation fidelity of the viral RNA-dependent RNA polymerase: a kinetic, thermodynamic and structural perspective. Virus Research *107*, 141-149.

Castro,C., Smidansky,E., Maksimchuk,K.R., Arnold,J.J., Korneeva,V.S., Gotte,M., Konigsberg,W., and Cameron,C.E. (2007). Two proton transfers in the transition state for nucleotidyl transfer catalyzed by RNA- and DNA-dependent RNA and DNA polymerases. Proceedings of the National Academy of Sciences of the United States of America *104*, 4267-4272.

Chambers,T.J., Grakoui,A., and Rice,C.M. (1991). Processing of the Yellow-Fever Virus Nonstructural Polyprotein - A Catalytically Active Ns3-Proteinase Domain and Ns2B Are Required for Cleavages at Dibasic Sites. Journal of Virology *65*, 6042-6050.

Chambers,T.J., Hahn,C.S., Galler,R., and Rice,C.M. (1990a). Flavivirus Genome Organization, Expression, and Replication. Annual Review of Microbiology *44*, 649-688.

Chambers,T.J., Mccourt,D.W., and Rice,C.M. (1990b). Production of yellow fever virus proteins in infected cells: identification of discrete polyprotein species and analysis of cleavage kinetics using region-specific polyclonal antisera. Virology *177*, 159-174.

Chambers,T.J., Nestorowicz,A., Amberg,S.M., and Rice,C.M. (1993). Mutagenesis of the Yellow-Fever Virus Ns2B Protein - Effects on Proteolytic Processing, Ns2B-Ns3 Complex-Formation, and Viral Replication. Journal of Virology *67*, 6797-6807.

Chambers,T.J. and Rice,C.M. (1987). Molecular-Biology of the Flaviviruses. Microbiological Sciences *4*, 219-223.

Chan,L., Das,S.K., Reddy,T.J., Poisson,C., Proulx,M., Pereira,O., Courchesne,M., Roy,C., Wang,W.Y., Siddiqui,A., Yannopoulos,C.G., Nguyen-Ba,N., Labrecque,D., Bethell,R., Hamel,M., Courtemanche-Asselin,P., L'Heureux,L., David,M., Nicolas,O., Brunette,S., Bilimoria,D., and Bedard,J. (2004). Discovery of thiophene-2-carboxylic acids as potent inhibitors of HCVNS5B polymerase and HCV subgenomic RNA replication. Part 1: Sulfonamides. Bioorganic & Medicinal Chemistry Letters *14*, 793-796.

Chang,Y.S., Liao,C.L., Tsao,C.H., Chen,M.C., Liu,G.I., Chen,L.K., and Lin,Y.L. (1999). Membrane permeabilization by small hydrophobic nonstructural proteins of Japanese encephalitis virus. Journal of Virology *73*, 6257-6264.

Chatterjee,D.K. and Esposito,D. (2006). Enhanced soluble protein expression using two new fusion tags. Protein Expression and Purification *46*, 122-129.

Chaturvedi,U.C., Shrivastava,R., and Nagar,R. (2005). Dengue vaccines: Problems & prospects. Indian Journal of Medical Research *121*, 639-652.

Chen,C.J., Kuo,M.D., Chien,L.J., Hsu,S.L., Wang,Y.M., and Lin,J.H. (1997). RNA-protein interactions: Involvement of NS3, NS5, and 3' noncoding regions of Japanese encephalitis virus genomic RNA. Journal of Virology *71*, 3466-3473.

Chen,J.B. and Ahlquist,P. (2000). Brome mosaic virus polymerase-like protein 2a is directed to the endoplasmic reticulum by helicase-like viral protein 1a. Journal of Virology *74*, 4310-4318.

Choi,K.H., Gallei,A., Becher,P., and Rossmann,M.G. (2006). The structure of bovine viral diarrhea virus RNA-dependent RNA polymerase and its amino-terminal domain. Structure *14*, 1107-1113.

Choi,K.H., Groarke,J.M., Young,D.C., Rossmann,M.G., Pevear,D.C., Kuhn,R.J., and Smith,J.L. (2004). Design, expression, and purification of a Flaviviridae polymerase using a high-throughput approach to facilitate crystal structure determination. Protein Science *13*, 2685-2692.

Christ,D. and Winter,G. (2006). Identification of protein domains by shotgun proteolysis. Journal of Molecular Biology *358*, 364-371.

Chu,P.W.G. and Westaway,E.G. (1985). Replication Strategy of Kunjin Virus - Evidence for Recycling Role of Replicative Form Rna As Template in Semiconservative and Asymmetric Replication. Virology *140*, 68-79.

Chu,P.W.G. and Westaway,E.G. (1987). Characterization of Kunjin Virus Rna-Dependent Rna-Polymerase - Reinitiation of Synthesis Invitro. Virology *157*, 330-337.

Clark,D.V., Mammen,M.P., Nisalak,A., Puthimethee,V., and Endy,T.P. (2005). Economic impact of dengue fever/dengue hemorrhagic fever in Thailand at the family and population levels. American Journal of Tropical Medicine and Hygiene *72*, 786-791.

Coia,G., Parker,M.D., Speight,G., Byrne,M.E., and Westaway,E.G. (1988). Nucleotide and Complete Amino-Acid Sequences of Kunjin Virus - Definitive Gene Order and Characteristics of the Virus-Specified Proteins. Journal of General Virology *69*, 1-21.

Corver,J., Lenches,E., Smith,K., Robison,R.A., Sando,T., Strauss,E.G., and Strauss,J.H. (2003). Fine mapping of a cis-acting sequence element in yellow fever virus RNA that is required for RNA replication and cyclization. Journal of Virology *77*, 2265-2270.

Crooks,A.J., Lee,J.M., Easterbrook,L.M., Timofeev,A.V., and Stephenson,J.R. (1994). The Ns1 Protein of Tick-Borne Encephalitis-Virus Forms Multimeric Species Upon Secretion from the Host-Cell. Journal of General Virology *75*, 3453-3460.

Crotty,S., Cameron,C.E., and Andino,R. (2001). RNA virus error catastrophe: Direct molecular test by using ribavirin. Proceedings of the National Academy of Sciences of the United States of America *98*, 6895-6900.

Cui,T.A., Sugrue,R.J., Xu,Q.R., Lee,A.K.W., Chan,Y.C., and Fu,J.L. (1998). Recombinant dengue virus type 1 NS3 protein exhibits specific viral RNA binding and NTPase activity regulated by the NS5 protein. Virology *246*, 409-417.

Cummings,M.D., Farnum,M.A., and Nelen,M.I. (2006). Universal screening methods and applications of ThermoFluor (R). Journal of Biomolecular Screening *11*, 854-863.

D'Arcy,A., Stihle,M., Kostrewa,D., and Dale,G. (1999). Crystal engineering: a case study using the 24 kDa fragment of the DNA gyrase B subunit from Escherichia coli. Acta Crystallographica Section D-Biological Crystallography *55*, 1623-1625.

Danley,D.E. (2006). Crystallization to obtain protein-ligand complexes for structure-aided drug design. Acta Crystallographica Section D-Biological Crystallography *62*, 569-575.

Declercq,E. (1995). Antiviral Therapy for Human-Immunodeficiency-Virus Infections. Clinical Microbiology Reviews *8*, 200-239.

Delarue,M., Boule,J.B., Lescar,J., Expert-Bezancon,N., Jourdan,N., Sukumar,N., Rougeon,F., and Papanicolaou,C. (2002). Crystal structures of a template-independent DNA polymerase: murine terminal deoxynucleotidyltransferase. Embo Journal *21*, 427-439.

Derewenda,Z.S. (2004). The use of recombinant methods and molecular engineering in protein crystallization. Methods *34*, 354-363.

DiTursi,M.K., Kwon,S.J., Reeder,P.J., and Dordick,J.S. (2006). Bioinformatics-driven, rational engineering of protein thermostability. Protein Engineering Design & Selection *19*, 517-524.

Dhanak,D., Duffy,K.J., Johnston,V.K., Lin-Goerke,J., Darey,M., Shaw,A.N., Gu,B.H., Silverman,C., Gates,A.T., Nonnemacher,M.R., Earnshaw,D.L., Casper,D.J., Kaura,A.,

Baker,A., Greenwood,C., Gutshall,L.L., Maley,D., DelVecchio,A., Macarron,R., Hofmann,G.A., Alnoah,Z., Cheng,H.Y., Chan,G., Khandekar,S., Keenan,R.M., and Sarisky,R.T. (2002). Identification and biological characterization of heterocyclic inhibitors of the hepatitis C virus RNA-dependent RNA polymerase. Journal of Biological Chemistry *277*, 38322-38327.

Dolja,V.V. and Koonin,E.V. (1991). Phylogeny of Capsid Proteins of Small Icosahedral Rna Plant-Viruses. Journal of General Virology *72*, 1481-1486.

Doublie,S., Tabor,S., Long,A.M., Richardson,C.C., and Ellenberger,T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 angstrom resolution. Nature *391*, 251-258.

Dreher,T.W. (1999). Functions of the 3 '-untranslated regions of positive strand RNA viral genomes. Annual Review of Phytopathology *37*, 151-174.

Egger,D., Pasamontes,L., Bolten,R., Boyko,V., and Bienz,K. (1996). Reversible dissociation of the poliovirus replication complex: Functions and interactions of its components in viral RNA synthesis. Journal of Virology *70*, 8675-8683.

Egger,D., Wolk,B., Gosert,R., Bianchi,L., Blum,H.E., Moradpour,D., and Bienz,K. (2002). Expression of hepatitis C virus proteins induces distinct membrane alterations including a candidate viral replication complex. Journal of Virology *76*, 5974-5984.

Egloff,M.P., Benarroch,D., Selisko,B., Romette,J.L., and Canard,B. (2002). An RNA cap (nucleoside-2 '-O-)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization. Embo Journal *21*, 2757-2768.

Eldrup,A.B., Allerson,C.R., Bennett,C.F., Bera,S., Bhat,B., Bhat,N., Bosserman,M.R., Brooks,J., Burlein,C., Carroll,S.S., Cook,P.D., Getty,K.L., MacCoss,M., McMasters,D.R., Olsen,D.B., Prakash,T.P., Prhavc,M., Song,Q.L., Tomassini,J.E., and Xia,J. (2004). Structure-activity relationship of purine ribonucleosides for inhibition of hepatitis C virus RNA-dependent RNA polymerase. Journal of Medicinal Chemistry *47*, 2283-2295.

Erbel,P., Schiering,N., D'Arcy,A., Renatus,M., Kroemer,M., Lim,S.P., Yin,Z., Keller,T.H., Vasudevan,S.G., and Hommel,U. (2006). Structural basis for the activation of flaviviral NS3 proteases from dengue and West Nile virus. Nature Structural & Molecular Biology *13*, 372-373.

Esnouf,R.M., Ren,J.S., Garman,E.F., Somers,D.O., Ross,C.K., Jones,E.Y., Stammers,D.K., and Stuart,D.I. (1998). Continuous and discontinuous changes in the unit cell of HIV-1 reverse transcriptase crystals on dehydration. Acta Crystallographica Section D-Biological Crystallography *54*, 938-953.

Esposito,D. and Chatterjee,D.K. (2006). Enhancement of soluble protein expression through the use of fusion tags. Current Opinion in Biotechnology *17*, 353-358.

Evans,E.J., Esnouf,R.M., Manso-Sancho,R., Gilbert,R.J.C., James,J.R., Yu,C., Fennelly,J.A., Vowles,C., Hanke,T., Walse,B., Hunig,T., Sorensen,P., Stuart,D.I., and Davis,S.J. (2005). Crystal structure of a soluble CD28-Fab complex. Nature Immunology *6*, 271-279.

Falconar,A.K.I. (1997). The dengue virus nonstructural-1 protein (NS1) generates antibodies to common epitopes on human blood clotting, integrin/adhesin proteins and binds to human endothelial cells: Potential implications in haemorrhagic fever pathogenesis. Archives of Virology *142*, 897-916.

Falgout,B., Chanock,R., and Lai,C.J. (1989). Proper processing of dengue virus nonstructural glycoprotein NS1 requires the N-terminal hydrophobic signal sequence and the downstream nonstructural protein NS2a. J. Virol. *63*, 1852-1860.

Falgout,B. and Markoff,L. (1995). Evidence that flavivirus NS1-NS2A cleavage is mediated by a membrane-bound host protease in the endoplasmic reticulum. J. Virol. *69*, 7232-7243.

Ferrari,E., Wright-Minogue,J., Fang,J.W.S., Baroudy,B.M., Lau,J.Y.N., and Hong,Z. (1999). Characterization of soluble hepatitis C virus RNA-dependent RNA polymerase expressed in Escherichia coli. Journal of Virology *73*, 1649-1654.

Ferrer-Orta,C., Arias,A., Agudo,R., Perez-Luque,R., Escarmis,C., Domingo,E., and Verdaguer,N. (2006a). The structure of a protein primer-polymerase complex in the initiation of genome replication. EMBO J. *25*, 880-888.

Ferrer-Orta,C., Arias,A., Escarmis,C., and Verdaguer,N. (2006b). A comparison of viral RNA-dependent RNA polymerases. Curr. Opin. Struct. Biol. *16*, 27-34.

Ferrer-Orta,C., Arias,A., Perez-Luque,R., Escarmis,C., Domingo,E., and Verdaguer,N. (2004). Structure of foot-and-mouth disease virus RNA-dependent RNA polymerase and its complex with a template-primer RNA. J. Biol. Chem. *279*, 47212-47221.

Ferrer-Orta,C., Arias,A., Perez-Luque,R., Escarmis,C., Domingo,E., and Verdaguer,N. (2007). Sequential structures provide insights into the fidelity of RNA replication. Proceedings of the National Academy of Sciences of the United States of America *104*, 9463-9468.

Filomatori,C.V., Lodeiro,M.F., Alvarez,D.E., Samsa,M.M., Pietrasanta,L., and Gamarnik,A.V. (2006). A 5 ' RNA element promotes dengue virus RNA synthesis on a circular genome. Genes & Development *20*, 2238-2249.

Flamand,M., Megret,F., Mathieu,M., Lepault,J., Rey,F.A., and Deubel,V. (1999). Dengue virus type 1 nonstructural glycoprotein NS1 is secreted from mammalian cells as a soluble hexamer in a glycosylation-dependent fashion. Journal of Virology *73*, 6104-6110.

Fontana,A., Fassina,G., Vita,C., Dalzoppo,D., Zamai,M., and Zambonin,M. (1986). Correlation Between Sites of Limited Proteolysis and Segmental Mobility in Thermolysin. Biochemistry *25*, 1847-1851.

Forwood,J.K., Brooks,A., Briggs,L.J., Xiao,C.Y., Jans,D.A., and Vasudevan,S.G. (1999). The 37-amino-acid interdomain of dengue virus NS5 protein contains a functional NLS and inhibitory CK2 site. Biochemical and Biophysical Research Communications *257*, 731-737.

Garcia-Montalvo,B.M., Medina,F., and del Angel,R.M. (2004). La protein binds to NS5 and NS3 and to the 5 ' and 3 ' ends of Dengue 4 virus RNA. Virus Research *102*, 141-150.

Gelpi,E., Preusser,M., Garzuly,F., Holzmann,H., Heinz,F.X., and Budka,H. (2005). Visualization of central European tick-borne encephalitis infection in fatal human cases. Journal of Neuropathology and Experimental Neurology *64*, 506-512.

Gohara,D.W., Arnold,J.J., and Cameron,C.E. (2004). Poliovirus RNA-dependent RNA polymerase (3D(pol)): Kinetic, thermodynamic, and structural analysis of ribonucleotide selection. Biochemistry *43*, 5149-5158.

Gohara,D.W., Crotty,S., Arnold,J.J., Yoder,J.D., Andino,R., and Cameron,C.E. (2000). Poliovirus RNA-dependent RNA polymerase (3D(pol)) - Structural, biochemical, and biological analysis of conserved structural motifs A and B. Journal of Biological Chemistry *275*, 25523-25532.

Goldschmidt,L., Cooper,D.R., Derewenda,Z.S., and Eisenberg,D. (2007). Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. Protein Science *16*, 1569-1576.

Goldsmith-Fischman,S. and Honig,B. (2003). Structural genomics: Computational methods for structure analysis. Protein Science *12*, 1813-1821.

Gorbalenya,A.E., Donchenko,A.P., Koonin,E.V., and Blinov,V.M. (1989). N-Terminal Domains of Putative Helicases of Flaviviruses and Pestiviruses May be Serine Proteases. Nucleic Acids Research *17*, 3889-3897.

Gorbalenya,A.E. and Koonin,E.V. (1993). Helicases - Amino-Acid-Sequence Comparisons and Structure-Function-Relationships. Current Opinion in Structural Biology *3*, 419-429.

Goulding,C.W. and Perry,L.J. (2003). Protein production in Escherichia coli for structural studies by X-ray crystallography. Journal of Structural Biology *142*, 133-143.

Grassmann,C.W., Isken,O., Tautz,N., and Behrens,S.E. (2001). Genetic analysis of the pestivirus nonstructural coding region: Defects in the NS5A unit can be complemented in trans. Journal of Virology *75*, 7791-7802.

Gubler,D.J. (1996). The global resurgence of arboviral diseases. Transactions of the Royal Society of Tropical Medicine and Hygiene *90*, 449-451.

Gubler,D.J. (1998). The global pandemic of dengue/dengue haemorrhagic fever: current status and prospects for the future. Ann. Acad. Med. Singapore *27*, 227-234.

Gubler,D.J. (2001). Human arbovirus infections worldwide. West Nile Virus: Detection, Surveillance, and Control *951*, 13-24.

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

Gubler,D.J. (2002b). The global emergence/resurgence of arboviral diseases as public health problems. Archives of Medical Research *33*, 330-342.

Gubler,D.J. and Meltzer,M. (1999). Impact of dengue/dengue hemorrhagic fever on the developing world. Advances in Virus Research, Vol 53 *53*, 35-70.

Gulick,R.M., Mellors,J.W., Havlir,D., Eron,J.J., Gonzalez,C., McMahon,D., Richman,D.D., Valentine,F.T., Jonas,L., Meibohm,A., Emini,E.A., and Chodakewitz,J.A. (1997). Treatment with indinavir, zidovudine, and lamivudine in adults with human immunodeficiency virus infection and prior antiretroviral therapy. New England Journal of Medicine *337*, 734-739.

Guzman,M.G. and Kouri,G. (2002). Dengue: an update. Lancet Infectious Diseases *2*, 33-42.

Hagen,M., Tiley,L., Chung,T.D.Y., and Krystal,M. (1995). The Role of Template-Primer Interactions in Cleavage and Initiation by the Influenza-Virus Polymerase. Journal of General Virology *76*, 603-611.

Hahn,C.S., Hahn,Y.S., Rice,C.M., Lee,E., Dalgarno,L., Strauss,E.G., and Strauss,J.H. (1987). Conserved Elements in the 3' Untranslated Region of Flavivirus Rnas and Potential Cyclization Sequences. Journal of Molecular Biology *198*, 33-41.

Hamdan,S.M., Johnson,D.E., Tanner,N.A., Lee,J.B., Qimron,U., Tabor,S., van Oijen,A.M., and Richardson,C.C. (2007). Dynamic DNA helicase-DNA polymerase interactions assure processive replication fork movement. Molecular Cell *27*, 539-549.

Hammon,W.M., RUDNICK,A., Sather,G., ROGERS,K.D., and MORSE,L.J. (1960). New hemorrhagic fevers of children in the Philippines and Thailand. Trans. Assoc. Am. Physicians *73*, 140-155.

Hanley,K.A., Lee,J.J., Blaney,J.E., Jr., Murphy,B.R., and Whitehead,S.S. (2002). Paired charge-to-alanine mutagenesis of dengue virus type 4 NS5 generates mutants with temperature-sensitive, host range, and mouse attenuation phenotypes. J. Virol. *76*, 525-531.

Hansen,J.L., Long,A.M., and Schultz,S.C. (1997). Structure of the RNA-dependent RNA polymerase of poliovirus. Structure *5*, 1109-1122.

Harris,K.S., Xiang,W.K., Alexander,L., Lane,W.S., Paul,A.V., and Wimmer,E. (1994). Interaction of Poliovirus Polypeptide 3Cd(Pro) with the 5'-Termini and 3'-Termini of the Poliovirus Genome - Identification of Viral and Cellular Cofactors Needed for Efficient Binding. Journal of Biological Chemistry *269*, 27004-27014.

Hassell,A.M., An,G., Bledsoe,R.K., Bynum,J.M., Carter,H.L., Deng,S.J.J., Gampe,R.T., Grisard,T.E., Madauss,K.P., Nolte,R.T., Rocque,W.J., Wang,L.P., Weaver,K.L., Williams,S.P., Wisely,G.B., Xu,R., and Shewchuk,L.M. (2007). Crystallization of protein-ligand complexes. Acta Crystallographica Section D-Biological Crystallography *63*, 72-79.

Hayes,R.J., Bentzien,J., Ary,M.L., Hwang,M.Y., Jacinto,J.M., Vielmetter,J., Kundu,A., and Dahiyat,B.I. (2002). Combining computational and experimental screening for rapid optimization of protein properties. Proceedings of the National Academy of Sciences of the United States of America *99*, 15926-15931.

Henchal,E.A. and Putnak,J.R. (1990). The dengue viruses. Clin. Microbiol. Rev. *3*, 376-396.

Heras,B. and Martin,J.L. (2005). Post-crystallization treatments for improving diffraction quality of protein crystals. Acta Crystallographica Section D-Biological Crystallography *61*, 1173-1180.

Holmes,E.C. and Twiddy,S.S. (2003). The origin, emergence and evolutionary genetics of dengue virus. Infect. Genet. Evol. *3*, 19-28.

Hong,Z., Cameron,C.E., Walker,M.P., Castro,C., Yao,N.H., Lau,J.Y.N., and Zhong,W.D. (2001). A novel mechanism to ensure terminal initiation by hepatitis C virus NS5B polymerase. Virology *285*, 6-11.

Hotta,S. (1952). Experimental studies on dengue. I. Isolation, identification and modification of the virus. J. Infect. Dis. *90*, 1-9.

Huang,H.F., Chopra,R., Verdine,G.L., and Harrison,S.C. (1998). Structure of a covalently trapped catalytic complex of HIV-I reverse transcriptase: Implications for drug resistance. Science *282*, 1669-1675.

Huang,S.G. (2003). Limited proteolysis reveals conformational changes in uncoupling protein-1 from brown adipose tissue mitochondria. Archives of Biochemistry and Biophysics *420*, 40-45.

Hubbard,S.J. (1998). The structural aspects of limited proteolysis of native proteins. Biochimica et Biophysica Acta-Protein Structure and Molecular Enzymology *1382*, 191-206.

Hung,S.L., Lee,P.L., Chen,H.W., Chen,L.K., Kao,C.L., and King,C.C. (1999). Analysis of the steps involved in dengue virus entry into host cells. Virology *257*, 156-167.

Indiani,C. and O'Donnell,M. (2006). The replication clamp-loading machine at work in the three domains of life. Nature Reviews Molecular Cell Biology *7*, 751-761.

Jacobomolina,A., Ding,J.P., Nanni,R.G., Clark,A.D., Lu,X.D., Tantillo,C., Williams,R.L., Kamer,G., Ferris,A.L., Clark,P., Hizi,A., Hughes,S.H., and Arnold,E. (1993). Crystal-Structure of Human-Immunodeficiency-Virus Type-1 Reverse-Transcriptase Complexed with Double-Stranded Dna at 3.0 Angstrom Resolution Shows Bent Dna. Proceedings of the National Academy of Sciences of the United States of America *90*, 6320-6324.

Jager,J., Smerdon,S.J., Wang,J.M., Boisvert,D.C., and Steitz,T.A. (1994). Comparison of 3 Different Crystal Forms Shows Hiv-1 Reverse-Transcriptase Displays An Internal Swivel Motion. Structure *2*, 869-876.

Jakubiec,A., Tournier,V., Drugeon,G., Pflieger,S., Camborde,L., Vinh,J., Hericourt,F., Redeker,V., and Jupin,I. (2006). Phosphorylation of viral RNA-dependent RNA polymerase and its role in replication of a plus-strand RNA virus. Journal of Biological Chemistry *281*, 21236-21249.

Jan,L.R., Yang,C.S., Trent,D.W., Falgout,B., and Lai,C.J. (1995). Processing of Japanese Encephalitis-Virus Nonstructural Proteins - Ns2B-Ns3 Complex and Heterologous Proteases. Journal of General Virology *76*, 573-580.

Jayaram,H., Estes,M.K., and Prasad,B.V.V. (2004). Emerging themes in rotavirus cell entry, genome organization, transcription and replication. Virus Research *101*, 67-81.

Joyce,C.M. and Steitz,T.A. (1995). Polymerase Structures and Function - Variations on A Theme. Journal of Bacteriology *177*, 6321-6329.

Kao,C.C., Singh,P., and Ecker,D.J. (2001). De novo initiation of viral RNA-Dependent RNA synthesis. Virology *287*, 251-260.

Kapoor,M., Zhang,L.W., Ramachandra,M., Kusukawa,J., Ebner,K.E., and Padmanabhan,R. (1995). Association Between Ns3 and Ns5 Proteins of Dengue Virus Type-2 in the Putative Rna Replicase Is Linked to Differential Phosphorylation of Ns5. Journal of Biological Chemistry *270*, 19100-19106.

Kautner,I., Robinson,M.J., and Kuhnle,U. (1997). Dengue virus infection: epidemiology, pathogenesis, clinical presentation, diagnosis, and prevention. J. Pediatr. *131*, 516-524.

Khasnis,A.A. and Nettleman,M.D. (2005). Global warming and infectious disease. Arch. Med. Res. *36*, 689-696.

Khromykh,A.A., Kenney,M.T., and Westaway,E.G. (1998). trans-Complementation of flavivirus RNA polymerase gene NS5 by using Kunjin virus replicon-expressing BHK cells. J. Virol. *72*, 7270-7279.

Khromykh,A.A., Kondratieva,N., Sgro,J.Y., Palmenberg,A., and Westaway,E.G. (2003). Significance in replication of the terminal nucleotides of the Flavivirus genome. Journal of Virology *77*, 10623-10629.

Khromykh,A.A., Meka,H., Guyatt,K.J., and Westaway,E.G. (2001a). Essential role of cyclization sequences in Flavivirus RNA replication. Journal of Virology *75*, 6719-6728.

Khromykh,A.A., Sedlak,P.L., and Westaway,E.G. (2000). cis- and trans-acting elements in flavivirus RNA replication. Journal of Virology *74*, 3253-3263.

Khromykh,A.A., Varnavski,A.N., Sedlak,P.L., and Westaway,E.G. (2001b). Coupling between replication and packaging of flavivirus RNA: Evidence derived from the use of DNA-based full-length cDNA clones of kunjin virus. Journal of Virology *75*, 4633-4640.

Kiefer,J.R., Mao,C., Braman,J.C., and Beese,L.S. (1998). Visualizing DNA replication in a catalytically active Bacillus DNA polymerase crystal. Nature *391*, 304-307.

Kim,J.L., Morgenstern,K.A., Griffith,J.P., Dwyer,M.D., Thomson,J.A., Murcko,M.A., Lin,C., and Caron,P.R. (1998). Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding. Structure *6*, 89-100.

Kim,J.L., Morgenstern,K.A., Lin,C., Fox,T., Dwyer,M.D., Landro,J.A., Chambers,S.P., Markland,W., Lepre,C.A., OMalley,E.T., Harbeson,S.L., Rice,C.M., Murcko,M.A., Caron,P.R., and Thomson,J.A. (1996). Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide. Cell *87*, 343-355.

Kim,S.J., Kim,J.H., Kim,Y.G., Lim,H.S., and Oh,J.W. (2004). Protein kinase C-related kinase 2 regulates hepatitis C virus RNA polymerase function by phosphorylation. Journal of Biological Chemistry *279*, 50031-50041.

Koch,U. and Narjes,F. (2007). Recent progress in the development of inhibitors of the hepatitis c virus RNA-Dependent RNA polymerase. Current Topics in Medicinal Chemistry *7*, 1302-1329.

Kohlstaedt,L.A., Wang,J., Friedman,J.M., Rice,P.A., and Steitz,T.A. (1992). Crystal-Structure at 3.5 Angstrom Resolution of Hiv-1 Reverse-Transcriptase Complexed with An Inhibitor. Science *256*, 1783-1790.

Koonin,E.V. (1991). The Phylogeny of Rna-Dependent Rna-Polymerases of Positive-Strand Rna Viruses. Journal of General Virology *72*, 2197-2206.

Koonin,E.V. (1993). Computer-Assisted Identification of A Putative Methyltransferase Domain in Ns5 Protein of Flaviviruses and Lambda-2 Protein of Reovirus. Journal of General Virology *74*, 733-740.

Kopek,B.G., Perkins,G., Miller,D.J., Ellisman,M.H., and Ahlquist,P. (2007). Three-dimensional analysis of a viral RNA replication complex reveals a virus-induced mini-organelle. Plos Biology *5*, 2022-2034.

Kost,T.A., Condreay,J.P., and Jarvis,D.L. (2005). Baculovirus as versatile vectors for protein expression in insect and mammalian cells. Nature Biotechnology *23*, 567-575.

Kovari,L.C., Momany,C., and Rossmann,M.G. (1995). The use of antibody fragments for crystallization and structure determinations. Structure *3*, 1291-1293.

Kuhn,R.J., Zhang,W., Rossmann,M.G., Pletnev,S.V., Corver,J., Lenches,E., Jones,C.T., Mukhopadhyay,S., Chipman,P.R., Strauss,E.G., Baker,T.S., and Strauss,J.H. (2002). Structure of dengue virus: Implications for flavivirus organization, maturation, and fusion. Cell *108*, 717-725.

Kummerer,B.M. and Rice,C.M. (2002). Mutations in the yellow fever virus nonstructural protein NS2A selectively block production of infectious particles. Journal of Virology *76*, 4773-4784.

Lai,M.M.C. (1998). Cellular factors in the transcription and replication of viral RNA genomes: A parallel to DNA-dependent RNA transcription. Virology *244*, 1-12.

Lai,V.C.H., Kao,C.C., Ferrari,E., Park,J., Uss,A.S., Wright-Minogue,J., Hong,Z., and Lau,J.Y.N. (1999). Mutational analysis of bovine viral diarrhea virus RNA-dependent RNA polymerase. Journal of Virology *73*, 10129-10136.

Lain,S., Riechmann,J.L., Martin,M.T., and Garcia,J.A. (1989). Homologous Potyvirus and Flavivirus Proteins Belonging to A Superfamily of Helicase-Like Proteins. Gene *82*, 357-362.

Laurila,M.R.L., Makeyev,E.V., and Bamford,D.H. (2002). Bacteriophage phi 6 RNA-dependent RNA polymerase - Molecular details of initiating nucleic acid synthesis without primer. Journal of Biological Chemistry *277*, 17117-17124.

Lee,E. and Lobigs,M. (2000). Substitutions at the putative receptor-binding site of an encephalitic flavivirus alter virulence and host cell tropism and reveal a role for glycosaminoglycans in entry. Journal of Virology *74*, 8867-8875.

Lee,E. and Lobigs,M. (2002). Mechanism of virulence attenuation of glycosaminoglycan-binding variants of Japanese encephalitis virus and Murray valley encephalitis virus. Journal of Virology *76*, 4901-4911.

Lehmann,E., Brueckner,F., and Cramer,P. (2007). Molecular basis of RNA-dependent RNA polymerase II activity. Nature *450*, 445-449.

Leitmeyer,K.C., Vaughn,D.W., Watts,D.M., Salas,R., de Chacon,I.V., Ramos,C., and Rico-Hesse,R. (1999). Dengue virus structural differences that correlate with pathogenesis. Journal of Virology *73*, 4738-4747.

Leslie,A.G.W., Powell,H.R., Winter,G., Svensson,O., Spruce,D., McSweeney,S., Love,D., Kinder,S., Duke,E., and Nave,C. (2002). Automation of the collection and processing of X-ray diffraction data - a generic approach. Acta Crystallographica Section D-Biological Crystallography *58*, 1924-1928.

Leung,D., Schroder,K., White,H., Fang,N.X., Stoermer,M.J., Abbenante,G., Martin,J.L., Young,P.R., and Fairlie,D.P. (2001). Activity of recombinant dengue 2 virus NS3 protease in the presence of a truncated NS2B co-factor, small peptide substrates, and inhibitors. Journal of Biological Chemistry *276*, 45762-45771.

Li,Y., Korolev,S., and Waksman,G. (1998). Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of Thermus aquaticus DNA polymerase I: structural basis for nucleotide incorporation. Embo Journal *17*, 7514-7525.

Lin,C., Amberg,S.M., Chambers,T.J., and Rice,C.M. (1993). Cleavage at A Novel Site in the Ns4A Region by the Yellow-Fever Virus Ns2B-3 Proteinase Is A Prerequisite for Processing at the Downstream 4A/4B Signalase Site. Journal of Virology *67*, 2327-2335.

Lin,C.F., Chiu,S.C., Hsiao,Y.L., Wan,S.W., Lei,H.Y., Shiau,A.L., Liu,H.S., Yeh,T.M., Chen,S.H., Liu,C.C., and Lin,Y.S. (2005). Expression of cytokine, chemokine, and adhesion molecules during endothelial cell activation induced by antibodies against dengue virus nonstructural protein 1. Journal of Immunology *174*, 395-403.

Lin,C.F., Lei,H.Y., Shiau,A.L., Liu,C.C., Liu,H.S., Yeh,T.M., Chen,S.H., and Lin,Y.S. (2003). Antibodies from dengue patient sera cross-react with endothelial cells and induce damage. Journal of Medical Virology *69*, 82-90.

Lin,C.F., Lei,H.Y., Shiau,A.L., Liu,H.S., Yeh,T.M., Chen,S.H., Liu,C.C., Chiu,S.C., and Lin,Y.S. (2002a). Endothelial cell apoptosis induced by antibodies against dengue virus nonstructural protein 1 via production of nitric oxide (vol 169, pg 657, 2002). Journal of Immunology *169*, 2215.

Lin,Y.L., Leib,H.Y., Lin,Y.S., Yeh,T.M., Chen,S.H., and Liu,H.S. (2002b). Heparin inhibits dengue-2 virus infection of five human liver cell lines. Antiviral Research *56*, 93-96.

Lindenbach,B.D. and Rice,C.M. (1997). trans-Complementation of yellow fever virus NS1 reveals a role in early RNA replication. Journal of Virology *71*, 9608-9617.

Lindenbach,B.D. and Rice,C.M. (1999). Genetic interaction of flavivirus nonstructural proteins NS1 and NS4A as a determinant of replicase function. Journal of Virology *73*, 4611-4621.

Lindenbach,B.D. and Rice,C.M. (2003). Molecular biology of flaviviruses. Flaviviruses: Structure, Replication and Evolution *59*, 23-+.

Lipinski,C.A. (2000). Drug-like properties and the causes of poor solubility and poor permeability. Journal of Pharmacological and Toxicological Methods *44*, 235-249.

Liu,W.J., Chen,H.B., and Khromykh,A.A. (2003). Molecular and functional analyses of Kunjin virus infectious cDNA clones demonstrate the essential roles for NS2A in virus assembly and for a nonconservative residue in NS3 in RNA replication. Journal of Virology *77*, 7804-7813.

Liu,W.J., Sedlak,P.L., Kondratieva,N., and Khromykh,A.A. (2002). Complementation analysis of the flavivirus Kunjin NS3 and NS5 proteins defines the minimal regions essential for formation of a replication complex and shows a requirement of NS3 in cis for virus assembly. Journal of Virology *76*, 10766-10775.

Liu,W.J., Wang,X.J., Mokhonov,V.V., Shi,P.Y., Randall,R., and Khromykh,A.A. (2005). Inhibition of interferon signaling by the New York 99 strain and Kunjin subtype of West Nile virus involves blockage of STAT1 and STAT2 activation by nonstructural proteins. Journal of Virology *79*, 1934-1942.

Lorenz,I.C., Allison,S.L., Heinz,F.X., and Helenius,A. (2002). Folding and dimerization of tick-borne encephalitis virus envelope proteins prM and E in the endoplasmic reticulum. Journal of Virology *76*, 5480-5491.

Love,R.A., Parge,H.E., Wickersham,J.A., Hostomsky,Z., Habuka,N., Moomaw,E.W., Adachi,T., and Hostomska,Z. (1996). The crystal structure of hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site. Cell *87*, 331-342.

Luo,D., Xu,T., Hunke,C., Gruber,G., Vasudevan,S.G., and Lescar,J. (2008). Crystal structure of the NS3 protease-helicase from dengue virus. J. Virol. *82*, 173-183.

Luo,G., Wang,M., Konigsberg,W.H., and Xie,X.S. (2007). Single-molecule and ensemble fluorescence assays for a functionally important conformational change in T7 DNA polymerase. Proceedings of the National Academy of Sciences of the United States of America *104*, 12610-12615.

Lusty,C.J. (1999). A gentle vapor-diffusion technique for cross-linking of protein crystals for cryocrystallography. Journal of Applied Crystallography *32*, 106-112.

Mackenzie,J.M., Jones,M.K., and Young,Y.R. (1996). Immunolocalization of the dengue virus nonstructural glycoprotein NS1 suggests a role in viral RNA replication. Virology *220*, 232-240.

Mackenzie,J.M., Khromykh,A.A., Jones,M.K., and Westaway,E.G. (1998). Subcellular localization and some biochemical properties of the flavivirus Kunjin nonstructural proteins NS2A and NS4A. Virology *245*, 203-215.

Mackenzie,J.S., Gubler,D.J., and Petersen,L.R. (2004). Emerging flaviviruses: the spread and resurgence of Japanese encephalitis, West Nile and dengue viruses. Nature Medicine *10*, S98-S109.

Majumdar,C., Abbotts,J., Broder,S., and Wilson,S.H. (1988). Studies on the Mechanism of Human Immunodeficiency Virus Reverse-Transcriptase - Steady-State Kinetics, Processivity, and Polynucleotide Inhibition. Journal of Biological Chemistry *263*, 15657-15665.

Malet,H., Egloff,M.P., Selisko,B., Butcher,R.E., Wright,P.J., Roberts,M., Gruez,A., Sulzenbacher,G., Vonrhein,C., Bricogne,G., Mackenzie,J.M., Khromykh,A.A., Davidson,A.D., and Canard,B. (2007). Crystal structure of the RNA polymerase domain of the West Nile Virus non-structural protein 5. Journal of Biological Chemistry *282*, 10678-10689.

Martinez-Barragan,J.D. and del Angel,R.M. (2001). Identification of a putative coreceptor on vero cells that participates in dengue 4 virus infection. Journal of Virology *75*, 7818-7827.

Mason,P.W. (1989). Maturation of Japanese encephalitis virus glycoproteins produced by infected mammalian and mosquito cells. Virology *169*, 354-364.

Mastrangelo,E., Bollati,M., Milani,M., Brisbarre,N., de Lamballerie,X., Coutard,B., Canard,B., Khromykh,A., and Bolognesi,M. (2006). Preliminary crystallographic characterization of an RNA helicase from Kunjin virus. Acta Crystallographica Section F-Structural Biology and Crystallization Communications *62*, 876-879.

Mcclure,W.R. (1985). Mechanism and Control of Transcription Initiation in Prokaryotes. Annual Review of Biochemistry *54*, 171-204.

Medin,C.L., Fitzgerald,K.A., and Rothman,A.L. (2005). Dengue virus nonstructural protein NS5 induces interleukin-8 transcription and secretion. Journal of Virology *79*, 11053-11061.

Men,R.H., Bray,M., Clark,D., Chanock,R.M., and Lai,C.J. (1996). Dengue type 4 virus mutants containing deletions in the 3' noncoding region of the RNA genome: Analysis of growth restriction in cell culture and altered viremia pattern and immunogenicity in rhesus monkeys. Journal of Virology *70*, 3930-3937.

Migliaccio,G., Tomassini,J.E., Carroll,S.S., Tomei,L., Altamura,S., Bhat,B., Bartholomew,L., Bosserman,M.R., Ceccacci,A., Colwell,L.F., Cortese,R., De Francesco,R., Eldrup,A.B., Getty,K.L., Hou,X.S., LaFemina,R.L., Ludmerer,S.W., MacCoss,M., McMasters,D.R., Stahlhut,M.W., Olsen,D.B., Hazuda,D.J., and Flores,O.A. (2003). Characterization of resistance to non-obligate chain-terminating ribonucleoside analogs that inhibit hepatitis C virus replication in vitro. Journal of Biological Chemistry *278*, 49164-49170.

Miller,S., Kastner,S., Krijnse-Locker,J., Buhler,S., and Bartenschlager,R. (2007). The non-structural protein 4A of dengue virus is an integral membrane protein inducing membrane alterations in a 2K-regulated manner. J. Biol. Chem. *282*, 8873-8882.

Miller,S., Sparacio,S., and Bartenschlager,R. (2006). Subcellular localization and membrane topology of the Dengue virus type 2 Non-structural protein 4B. J. Biol. Chem. *281*, 8854-8863.

Modis,Y., Ogata,S., Clements,D., and Harrison,S.C. (2003). A ligand-binding pocket in the dengue virus envelope glycoprotein. Proceedings of the National Academy of Sciences of the United States of America *100*, 6986-6991.

Munoz-Jordan,J.L., Laurent-Rolle,M., Ashour,J., Martinez-Sobrido,L., Ashok,M., Lipkin,W.I., and Garcia-Sastre,A. (2005). Inhibition of alpha/beta interferon signaling by the NS4B protein of flaviviruses. Journal of Virology *79*, 8004-8013.

Munoz-Jordan,J.L., Sanchez-Burgos,G.G., Laurent-Rolle,M., and Garcia-Sastre,A. (2003). Inhibition of interferon signaling by dengue virus. Proceedings of the National Academy of Sciences of the United States of America *100*, 14333-14338.

Murphy,J.R. (1996). Protein engineering and design for drug delivery. Current Opinion in Structural Biology *6*, 541-545.

Murray,C.W., Callaghan,O., Chessari,G., Cleasby,A., Congreve,M., Frederickson,M., Hartshorn,M.J., McMenamin,R., Patel,S., and Wallis,N. (2007). Application of fragment screening by X-ray crystallography to beta-secretase. Journal of Medicinal Chemistry *50*, 1116-1123.

Navarro-Sanchez,E., Altmeyer,R., Amara,A., Schwartz,O., Fieschi,F., Virelizier,J.L., Arenzana-Seisdedos,F., and Despres,P. (2003). Dendritic-cell-specific ICAM3-grabbing non-integrin is essential for the productive infection of human dendritic cells by mosquito-cell-derived dengue viruses. Embo Reports *4*, 723-728.

Neyts,J. and Declercq,E. (1994). Mechanism of Action of Acyclic Nucleoside Phosphonates Against Herpes-Virus Replication. Biochemical Pharmacology *47*, 39-41.

Nestorowicz,A., Chambers,T.J., and Rice,C.M. (1994). Mutagenesis of the Yellow-Fever Virus Ns2A/2B Cleavage Site - Effects on Proteolytic Processing, Viral Replication, and Evidence for Alternative Processing of the Ns2A Protein. Virology *199*, 114-123.

Ng,C.Y., Gu,F., Phong,W.Y., Chen,Y.L., Lim,S.P., Davidson,A., and Vasudevan,S.G. (2007). Construction and characterization of a stable subgenomic dengue virus type 2 replicon system for antiviral compound and siRNA testing. Antiviral Research *76*, 222-231.

Ng,K.K., Cherney,M.M., Vazquez,A.L., Machin,A., Alonso,J.M., Parra,F., and James,M.N. (2002). Crystal structures of active and inactive conformations of a caliciviral RNA-dependent RNA polymerase. J. Biol. Chem. *277*, 1381-1387.

Ng,K.K., Pendas-Franco,N., Rojo,J., Boga,J.A., Machin,A., Alonso,J.M., and Parra,F. (2004). Crystal structure of norwalk virus polymerase reveals the carboxyl terminus in the active site cleft. J. Biol. Chem. *279*, 16638-16645.

Nomaguchi,M., Ackermann,M., Yon,C., You,S., and Padmanabhan,R. (2003). De novo synthesis of negative-strand RNA by Dengue virus RNA-dependent RNA polymerase in vitro: nucleotide, primer, and template parameters. J. Virol. *77*, 8831-8842.

Nova-Ocampo,M., Villegas-Sepulveda,N., and del Angel,R.M. (2002). Translation elongation factor-1alpha, La, and PTB interact with the 3' untranslated region of dengue 4 virus RNA. Virology *295*, 337-347.

Novotny,J. and Bruccoleri,R.E. (1987). Correlation Among Sites of Limited Proteolysis, Enzyme Accessibility and Segmental Mobility. Febs Letters *211*, 185-189.

O'Farrell,D., Trowbridge,R., Rowlands,D., and Jager,J. (2003). Substrate complexes of hepatitis C virus RNA polymerase (HC-J4): structural evidence for nucleotide import and de-novo initiation. J. Mol. Biol. *326*, 1025-1035.

O'Reilly,E.K. and Kao,C.C. (1998). Analysis of RNA-dependent RNA polymerase structure and function as guided by known polymerase structures and computer predictions of secondary structure. Virology *252*, 287-303.

Ollis,D.L., Brick,P., Hamlin,R., Xuong,N.G., and Steitz,T.A. (1985). Structure of Large Fragment of Escherichia-Coli Dna-Polymerase-I Complexed with Dtmp. Nature *313*, 762-766.

Ortin,J. and Parra,F. (2006). Structure and function of RNA replication. Annual Review of Microbiology *60*, 305-326.

Owen,R.L. and Garman,E. (2005). A new method for predetermining the diffraction quality of protein crystals: using SOAP as a selection tool. Acta Crystallographica Section D-Biological Crystallography *61*, 130-140.

Patton,J.T., Carpi,R.V.D., and Spencer,E. (2004). Replication and transcription of the rotavirus genome. Current Pharmaceutical Design *10*, 3769-3777.

Pechik,I., Nachman,J., Ingham,K., and Gilliland,G.L. (1993). Crystallization and Preliminary-X-Ray Diffraction Data of 2 Heparin-Binding Fragments of Human Fibronectin. Proteins-Structure Function and Genetics *16*, 43-47.

Perrone,P., Daverio,F., Valente,R., Rajyaguru,S., Martin,J.A., Leveque,V., Le Pogam,S., Najera,I., Klumpp,K., Smith,D.B., and McGuigan,C. (2007). First example of phosphoramidate approach applied to a 4 '-substituted purine nucleoside (4 '-azidoadenosine): Conversion of an inactive nucleoside to a submicromolar compound versus hepatitis C virus. Journal of Medicinal Chemistry *50*, 5463-5470.

Pierra,C., Amador,A., Benzaria,S., Cretton-Scott,E., D'Amours,M., Mao,J., Mathieu,S., Moussa,A., Bridges,E.G., Standring,D.N., Sommadossi,J.P., Storer,R., and Gosselin,G. (2006). Synthesis and pharmacokinetics of valopicitabine (NM283), an efficient prodrug of

the potent anti-HCV agent 2 '-C-methylcytidine. Journal of Medicinal Chemistry *49*, 6614-6620.

Poch,O., Sauvaget,I., Delarue,M., and Tordo,N. (1989). Identification of 4 Conserved Motifs Among the Rna-Dependent Polymerase Encoding Elements. Embo Journal *8*, 3867-3874.

Post,P.R., Carvalho,R., and Galler,R. (1991). Glycosylation and secretion of yellow fever virus nonstructural protein NS1. Virus Res. *18*, 291-302.

Prongay,A.J., Smith,T.J., Rossmann,M.G., Ehrlich,L.S., Carter,C.A., and Mcclure,J. (1990). Preparation and Crystallization of A Human-Immunodeficiency-Virus P24-Fab Complex. Proceedings of the National Academy of Sciences of the United States of America *87*, 9980-9984.

Pryor,M.J., Rawlinson,S.M., Butcher,R.E., Barton,C.L., Waterhouse,T.A., Vasudevan,S.G., Bardin,P.G., Wright,P.J., Jans,D.A., and Davidson,A.D. (2007). Nuclear localization of dengue virus nonstructural protein 5 through its importin alpha/beta-recognized nuclear localization sequences is integral to viral infection. Traffic *8*, 795-807.

Qu,L., McMullan,L.K., and Rice,C.M. (2001). Isolation and characterization of noncytopathic pestivirus mutants reveals a role for nonstructural protein NS4B in viral cytopathogenicity. Journal of Virology *75*, 10651-10662.

Quadt,R., Kao,C.C., Browning,K.S., Hershberger,R.P., and Ahlquist,P. (1993). Characterization of A Host Protein Associated with Brome Mosaic-Virus Rna-Dependent Rna-Polymerase. Proceedings of the National Academy of Sciences of the United States of America *90*, 1498-1502.

Ranjith-Kumar,C.T., Gutshall,L., Sarisky,R.T., and Kao,C.C. (2003). Multiple interactions within the hepatitis C virus RNA polymerase repress primer-dependent RNA synthesis. Journal of Molecular Biology *330*, 675-685.

Ranjith-Kumar,C.T., Kim,Y.C., Gutshall,L., Silverman,C., Khandekar,S., Sarisky,R.T., and Kao,C.C. (2002). Mechanism of de novo initiation by the hepatitis C virus RNA-dependent RNA polymerase: Role of divalent metals. Journal of Virology *76*, 12513-12525.

Ray,D., Shah,A., Tilgner,M., Guo,Y., Zhao,Y., Dong,H., Deas,T.S., Zhou,Y., Li,H., and Shi,P.Y. (2006). West Nile virus 5'-cap structure is formed by sequential guanine N-7 and ribose 2'-O methylations by nonstructural protein 5. J. Virol. *80*, 8362-8370.

Reardon,J.E. (1992). Human-Immunodeficiency-Virus Reverse-Transcriptase - Steady-State and Pre-Steady-State Kinetics of Nucleotide Incorporation. Biochemistry *31*, 4473-4479.

Reddy,T.J., Chan,L., Turcotte,N., Proulx,M., Pereira,O.Z., Das,S.K., Siddiqui,A., Wang,W.Y., Poisson,C., Yannopoulos,C.G., Bilimoria,D., L'Heureux,L., Alaoui,H.M.A., and Nguyen-Ba,N. (2003). Further SAR studies on novel small molecule inhibitors of the hepatitis C (HCV) NS5B polymerase. Bioorganic & Medicinal Chemistry Letters *13*, 3341-3344.

Reed,K.E., Gorbalenya,A.E., and Rice,C.M. (1998). The NS5A/NS5 proteins of viruses from three genera of the family Flaviviridae are phosphorylated by associated serine/threonine kinases. Journal of Virology *72*, 6199-6206.

Ren,J., Bird,L.E., Chamberlain,P.P., Stewart-Jones,G.B., Stuart,D.I., and Stammers,D.K. (2002). Structure of HIV-2 reverse transcriptase at 2.35-angstrom resolution and the mechanism of resistance to non-nucleoside inhibitors. Proceedings of the National Academy of Sciences of the United States of America *99*, 14410-14415.

Rey,F.A., Heinz,F.X., Mandl,C., Kunz,C., and Harrison,S.C. (1995). The Envelope Glycoprotein from Tick-Borne Encephalitis-Virus at 2 Angstrom Resolution. Nature *375*, 291-298.

Rieskautt,M.M. and Ducruix,A.F. (1989). Relative Effectiveness of Various Ions on the Solubility and Crystal-Growth of Lysozyme. Journal of Biological Chemistry *264*, 745-748.

Ruigrok,R.W.H. and Baudin,F. (1995). Structure of Influenza-Virus Ribonucleoprotein-Particles .2. Purified Rna-Free Influenza-Virus Ribonucleoprotein Forms Structures That Are Indistinguishable from the Intact Influenza-Virus Ribonucleoprotein-Particles. Journal of General Virology *76*, 1009-1014.

Sabin,A.B. (1952). Research on dengue during World War II. Am. J. Trop. Med. Hyg. *1*, 30-50.

Salgado,P.S., Makeyev,E.V., Butcher,S.J., Bamford,D.H., Stuart,D.I., and Grimes,J.M. (2004). The structural basis for RNA specificity and Ca2+ inhibition of an RNA-dependent RNA polymerase. Structure *12*, 307-316.

Sampath,A., Xu,T., Chao,A., Luo,D.H., Lescar,J., and Vasudevan,S.G. (2006). Structure-based mutational analysis of the NS3 helicase from dengue virus. Journal of Virology *80*, 6686-6690.

Sarafianos,S.G., Das,K., Clark,A.D., Ding,J.P., Boyer,P.L., Hughes,S.H., and Arnold,E. (1999). Lamivudine (3TC) resistance in HIV-1 reverse transcriptase involves steric hindrance with beta-branched amino acids. Proceedings of the National Academy of Sciences of the United States of America *96*, 10027-10032.

Sarafianos,S.G., Clark,A.D., Das,K., Tuske,S., Birktoft,J.J., Ilankumaran,P., Ramesha,A.R., Sayer,J.M., Jerina,D.M., Boyer,P.L., Hughes,S.H., and Arnold,E. (2002). Structures of HIV-1 reverse transcriptase with pre- and post-translocation AZTMP-terminated DNA. Embo Journal *21*, 6614-6624.

Schlesinger,J.J., Foltzer,M., and Chapman,S. (1993). The Fc Portion of Antibody to Yellow-Fever Virus-Ns1 Is A Determinant of Protection Against Yf Encephalitis in Mice. Virology *192*, 132-141.

Schul,W., Liu,W., Xu,H.Y., Flamand,M., and Vasudevan,S.G. (2007). A dengue fever viremia model in mice shows reduction in viral replication and suppression of the inflammatory response after treatment with antiviral drugs. Journal of Infectious Diseases *195*, 665-674.

Schwartz,M., Chen,J.B., Lee,W.M., Janda,M., and Ahlquist,P. (2004). Alternate, virus-induced membrane rearrangements support positive-strand RNA virus genome replication. Proceedings of the National Academy of Sciences of the United States of America *101*, 11263-11268.

Selisko,B., Dutartre,H., Guillemot,J.C., Debarnot,C., Benarroch,D., Khromykh,A., Despres,P., Egloff,M.P., and Canard,B. (2006). Comparative mechanistic studies of de novo RNA synthesis by flavivirus RNA-dependent RNA polymerases. Virology *351*, 145-158.

Shimakami,T., Honda,M., Kusakawa,T., Murata,T., Shimotohno,K., Kaneko,S., and Murakami,S. (2006). Effect of hepatitis C virus (HCV) NS5B-nucleolin interaction on HCV replication with HCV subgenomic replicon. Journal of Virology *80*, 3332-3340.

Singh,S.M. and Panda,A.K. (2005). Solubilization and refolding of bacterial inclusion body proteins. Journal of Bioscience and Bioengineering *99*, 303-310.

Smyth,D.R., Mrozkiewicz,M.K., McGrath,W.J., Listwan,P., and Kobe,B. (2003). Crystal structures of fusion proteins with large-affinity tags. Protein Science *12*, 1313-1322.

Sousa,R., Rose,J., and Wang,B.C. (1994). The Thumbs Knuckle - Flexibility in the Thumb Subdomain of T7 Rna-Polymerase Is Revealed by the Structure of A Chimeric T7/T3 Rna-Polymerase. Journal of Molecular Biology *244*, 6-12.

Staub,E., Fiziev,P., Rosenthal,A., and Hinzmann,B. (2004). Insights into the evolution of the nucleolus by an analysis of its protein domain repertoire. Bioessays *26*, 567-581.

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

Steitz,T.A. (1998). Structural biology - A mechanism for all polymerases. Nature *391*, 231-232.

Steitz,T.A. (1999). DNA polymerases: Structural diversity and common mechanisms. Journal of Biological Chemistry *274*, 17395-17398.

Steitz,T.A. (2006). Visualizing polynucleotide polymerase machines at work. Embo Journal *25*, 3458-3468.

Stemmer,W.P.C. (1994a). Dna Shuffling by Random Fragmentation and Reassembly - In-Vitro Recombination for Molecular Evolution. Proceedings of the National Academy of Sciences of the United States of America *91*, 10747-10751.

Stemmer,W.P.C. (1994b). Rapid Evolution of A Protein In-Vitro by Dna Shuffling. Nature *370*, 389-391.

Stiasny,K. and Heinz,F.X. (2006). Flavivirus membrane fusion. Journal of General Virology *87*, 2755-2766.

Svergun,D.I. and Koch,M.H.J. (2002). Advances in structure analysis using small-angle scattering in solution. Current Opinion in Structural Biology *12*, 654-660.

Tan,B.H., Fu,J., Sugrue,R.J., Yap,E.H., Chan,Y.C., and Tan,Y.H. (1996). Recombinant dengue type 1 virus NS5 protein expressed in Escherichia coli exhibits RNA-dependent RNA polymerase activity. Virology *216*, 317-325.

Tanner,N.K., Cordin,O., Banroques,J., Doere,M., and Linder,P. (2003). The Q motif: A newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis. Molecular Cell *11*, 127-138.

Tabor,S., Huber,H.E., and Richardson,C.C. (1987). Escherichia-Coli Thioredoxin Confers Processivity on the Dna-Polymerase-Activity of the Gene-5 Protein of Bacteriophage-T7. Journal of Biological Chemistry *262*, 16212-16223.

Tao,Y.Z., Farsetta,D.L., Nibert,M.L., and Harrison,S.C. (2002). RNA synthesis in a cage - Structural studies of reovirus polymerase lambda 3. Cell *111*, 733-745.

Tarbouriech,N., Curran,J., Ruigrok,R.W.H., and Burmeister,W.P. (2000). Tetrameric coiled coil domain of Sandal virus phosphoprotein. Nature Structural Biology *7*, 777-781.

Tedesco,R., Shaw,A.N., Bambal,R., Chai,D.P., Concha,N.O., Darcy,M.G., Dhanak,D., Fitch,D.M., Gates,A., Gerhardt,W.G., Halegoua,D.L., Han,C., Hofmann,G.A., Johnston,V.K., Kaura,A.C., Liu,N.N., Keenan,R.M., Lin-Goerke,J., Sarisky,R.T., Wiggall,K.J., Zimmerman,M.N., and Duffy,K.J. (2006). 3-(1,1-Dioxo-2H-(1,2,4)-benzothiadiazin-3-yl)-4-hydroxy-2(1H)-quinolinones, potent inhibitors of hepatitis C virus RNA-dependent RNA polymerase. Journal of Medicinal Chemistry *49*, 971-983.

Thompson,A.A. and Peersen,O.B. (2004). Structural basis for proteolysis-dependent activation of the poliovirus RNA-dependent RNA polymerase. EMBO J. *23*, 3462-3471.

Timofeev,A.V., Ozherelkov,S.V., Pronin,A.V., Deeva,A.V., Karganova,G.G., Elbert,L.B., and Stephenson,J.R. (1998). Immunological basis for protection in a murine model of tick-borne encephalitis by a recombinant adenovirus carrying the gene encoding the NS1 non-structural protein. Journal of General Virology *79*, 689-695.

Tomei,L., Altamura,S., Bartholomew,L., Biroccio,A., Ceccacci,A., Pacini,L., Narjes,F., Gennari,N., Bisbocci,M., Incitti,I., Orsatti,L., Harper,S., Stansfield,I., Rowley,M., De Francesco,R., and Migliaccio,G. (2003). Mechanism of action and antiviral activity of benzimidazole-based allosteric inhibitors of the hepatitis C virus RNA-dependent RNA polymerase. Journal of Virology *77*, 13225-13231.

Tomei,L., Altamura,S., Bartholomew,L., Bisbocci,M., Bailey,C., Bosserman,M., Cellucci,A., Forte,E., Incitti,I., Orsatti,L., Koch,U., De Francesco,R., Olsen,D.B., Carroll,S.S., and Migliaccio,G. (2004). Characterization of the inhibition of hepatitis C virus RNA replication by nonnucleosides. Journal of Virology *78*, 938-946.

Twiddy,S.S., Holmes,E.C., and Rambaut,A. (2003). Inferring the rate and time-scale of dengue virus evolution. Mol. Biol. Evol. *20*, 122-129.

Uchil,P.D., Kumar,A.V.A., and Satchidanandam,V. (2006). Nuclear localization of flavivirus RNA synthesis in infected cells. Journal of Virology *80*, 5451-5464.

Umareddy,I., Chao,A., Sampath,A., Gu,F., and Vasudevan,S.G. (2006). Dengue virus NS4B interacts with NS3 and dissociates it from single-stranded RNA. J. Gen. Virol. *87*, 2605-2614.

van Dijk,A.A., Makeyevt,E.V., and Bamford,D.H. (2004). Initiation of viral RNA-dependent RNA polymerization. Journal of General Virology *85*, 1077-1093.

Vasilyeva,O.V., Martynova,N.Y., Potapenko,N.A., and Ovchinnikova,T.V. (2004). Isolation and characterization of fragments of the ATP-dependent protease lon from Escherichia coli obtained by limited proteolysis. Russian Journal of Bioorganic Chemistry *30*, 306-314.

Venturi,M. and Hunte,C. (2003). Monoclonal antibodies for the structural analysis of the Na+/H+ antiporter NhaA from Escherichia coli. Biochimica et Biophysica Acta-Biomembranes *1610*, 46-50.

Wang,C.F., Gale,M., Keller,B.C., Huang,H., Brown,M.S., Goldstein,J.L., and Ye,J. (2005). Identification of FBL2 as a geranylgeranylated required for hepatitis C cellular protein virus RNA replication. Molecular Cell *18*, 425-434.

Wang,M., Ng,K.K., Cherney,M.M., Chan,L., Yannopoulos,C.G., Bedard,J., Morin,N., Nguyen-Ba,N., Alaoui-Ismaili,M.H., Bethell,R.C., and James,M.N. (2003). Non-nucleoside analogue inhibitors bind to an allosteric site on HCV NS5B polymerase. Crystal structures and mechanism of inhibition. J. Biol. Chem. *278*, 9489-9495.

Watashi,K., Ishii,N., Hijikata,M., Inoue,D., Murata,T., Miyanari,Y., and Shimotohno,K. (2005). Cyclophilin B is a functional regulator of hepatitis C virus RNA polymerase. Molecular Cell *19*, 111-122.

Wengler,G. and Wengler,G. (1993). The Ns-3 Nonstructural Protein of Flaviviruses Contains An Rna Triphosphatase-Activity. Virology *197*, 265-273.

Westaway,E.G., Khromykh,A.A., Kenney,M.T., Mackenzie,J.M., and Jones,M.K. (1997a). Proteins C and NS4B of the flavivirus Kunjin translocate independently into the nucleus. Virology *234*, 31-41.

Westaway,E.G., Mackenzie,J.M., Kenney,M.T., Jones,M.K., and Khromykh,A.A. (1997b). Ultrastructure of Kunjin virus-infected cells: Colocalization of NS1 and NS3 with double-stranded RNA, and of NS2B with NS3, in virus-induced membrane structures. Journal of Virology *71*, 6650-6661.

Westaway,E.G., Mackenzie,J.M., and Khromykh,A.A. (2003). Kunjin RNA replication and applications of Kunjin replicons. Flaviviruses: Structure, Replication and Evolution *59*, 99-+.

Winkler,G., Maxwell,S.E., Ruemmler,C., and Stollar,V. (1989). Newly Synthesized Dengue-2 Virus Nonstructural Protein Ns1 Is A Soluble-Protein But Becomes Partially Hydrophobic and Membrane-Associated After Dimerization. Virology *171*, 302-305.

Winkler,G., Randolph,V.B., Cleaves,G.R., Ryan,T.E., and Stollar,V. (1988). Evidence that the mature form of the flavivirus nonstructural protein NS1 is a dimer. Virology *162*, 187-196.

Wong,I., Patel,S.S., and Johnson,K.A. (1991). An Induced-Fit Kinetic Mechanism for Dna-Replication Fidelity - Direct Measurement by Single-Turnover Kinetics. Biochemistry *30*, 526-537.

Wu,C.F., Wang,S.H., Sun,C.M., Hu,S.T., and Syu,W.J. (2003). Activation of dengue protease autocleavage at the NS2B-NS3 junction by recombinant NS3 and GST-NS2B fusion proteins. Journal of Virological Methods *114*, 45-54.

Wu,J.H., Bera,A.K., Kuhn,R.J., and Smith,J.L. (2005a). Structure of the flavivirus helicase: Implications for catalytic activity, protein interactions, and proteolytic processing. Journal of Virology *79*, 10268-10277.

Wu,J.Z., Yao,N.H., Walker,M., and Hong,Z. (2005b). Recent advances in discovery and development of promising therapeutics against hepatitis C virus NS5B RNA-dependent RNA polymerase. Mini-Reviews in Medicinal Chemistry *5*, 1103-1112.

Wuite,G.J.L., Smith,S.B., Young,M., Keller,D., and Bustamante,C. (2000). Single-molecule studies of the effect of template tension on T7 DNA polymerase activity. Nature *404*, 103-106.

Xu,T., Sampath,A., Chao,A., Wen,D., Nanao,M., Chene,P., Vasudevan,S.G., and Lescar,J. (2005). Structure of the Dengue virus helicase/nucleoside triphosphatase catalytic domain at a resolution of 2.4 A. J. Virol. *79*, 10278-10288.

Yamshchikov,V.F. and Compans,R.W. (1994). Processing of the Intracellular Form of the West Nile Virus Capsid Protein by the Viral Ns2B-Ns3 Protease - An In-Vitro Study. Journal of Virology *68*, 5765-5771.

Yap,T.L., Xu,T., Chen,Y.L., Malet,H., Egloff,M.P., Canard,B., Vasudevan,S.G., and Lescar,J. (2007). Crystal Structure of the Dengue Virus RNA-Dependent RNA Polymerase Catalytic Domain at 1.85-Angstrom Resolution. J. Virol. *81*, 4753-4765.

You,S., Falgout,B., Markoff,L., and Padmanabhan,R. (2001a). In vitro RNA synthesis from exogenous dengue viral RNA templates requires long range interactions between 5'- and 3'-terminal regions that influence RNA structure. J. Biol. Chem. *276*, 15581-15591.

You,S. and Padmanabhan,R. (1999). A novel in vitro replication system for dengue virus - Initiation of RNA synthesis at the 3 '-end of exogenous viral RNA templates requires 5 '- and 3 '-terminal complementary sequence motifs of the viral RNA. Journal of Biological Chemistry *274*, 33714-33722.

You,S.Y., Falgout,B., Markoff,L., and Padmanabhan,R. (2001b). In vitro RNA synthesis from exogenous dengue viral RNA templates requires long range interactions between 5 '- and 3 '-terminal regions that influence RNA structure. Journal of Biological Chemistry *276*, 15581-15591.

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 virus protein NS1 in the sera of infected patients. Journal of Clinical Microbiology *38*, 1053-1057.

Yusof,R., Clum,S., Wetzel,M., Murthy,H.M.K., and Padmanabhan,R. (2000). Purified NS2B/NS3 serine protease of dengue virus type 2 exhibits cofactor NS2B dependence for cleavage of substrates with dibasic amino acids in vitro. Journal of Biological Chemistry *275*, 9963-9969.

Zamyatkin,D.F., Parra,F., Alonso,J.M.M., Harki,D.A., Peterson,B.R., Grochulski,P., and Ng,K.K.S. (2008). Structural insights into mechanisms of catalysis and inhibition in Norwalk virus polymerase. Journal of Biological Chemistry *283*, 7705-7712.

Zanotto,P.M., Gould,E.A., Gao,G.F., Harvey,P.H., and Holmes,E.C. (1996). Population dynamics of flaviviruses revealed by molecular phylogenies. Proc. Natl. Acad. Sci. U. S. A *93*, 548-553.

Zeng,L.L., Falgout,B., and Markoff,L. (1998). Identification of specific nucleotide sequences within the conserved 3 '-SL in the dengue type 2 virus genome required for replication. Journal of Virology *72* , 7510-7522.

Zhong,J., Gastaminza,P., Cheng,G.F., Kapadia,S., Kato,T., Burton,D.R., Wieland,S.F., Uprichard,S.L., Wakita,T., and Chisari,F.V. (2005). Robust hepatitis C virus infection in vitro. Proceedings of the National Academy of Sciences of the United States of America *102*, 9294-9299.

Appendix List A1

DEN3 RdRp protein expression and purification

8-10L preparation for crystal growth.

Expression

DNA fragments encoding the catalytic domains of the DENV RdRp from the four serotypes (DENV 1-4) spanning residues 273-900, were amplified by PCR and cloned into pET 15b (Novagen, Madison, USA), using the forward and reverse primers. First, the *E. coli* BL21-DE3 cells (RIL) (Stratagene) was transformed with plasmid with heat shock (42°C) and plated on LB agar containing 100 μg ml^{-1} ampicillin and 50 μg ml^{-1} chloramphenicol and grown at 37°C for overnight. One plate was prepared for each 500ml culture. The next day, *E. coli* colonies on the plate were washed with LB medium and inoculated into 500ml LB medium containing 100 μg ml^{-1} ampicillin and 50 μg ml^{-1} chloramphenicol and grown at 37°C until an OD_{600} of 0.4-0.6. The culture was cooled down at 4°C before protein expression was induced at 16°C by adding isopropyl-β-D-thiogalactopyranoside to a final concentration of 0.4 mM. After overnight growth, cells were harvested by centrifugation at 8,000 × *g* for 10 min at 277 K. Pellets were resuspended in buffer A (20 mM Tris-HCl pH 7.5, 0.5 M NaCl 10 mM β-mercaptoethanol, 10% Glycerol) supplemented with a protease inhibitor EDTA-free tablet (Roche, Switzerland). This was followed by sonication and centrifugation at 30,000 × g for 30 min at 277 K.

Purification

The lysate supernatants were first purified by metal affinity using a HisTrap HP column (GE Healthcare, Sweden) equilibrated with buffer A. Unbound proteins were washed sequentially with five column volumes of buffer A, supplemented with imidazole at concentrations of 25

mM and 125 mM respectively. Proteins were eluted using a linear gradient of imidazole ranging from 125 to 500 mM. Fractions containing the protein were pooled and dialyzed overnight against a) 50 mM Mes pH 6.8, 0.3 M NaCl, 1 mM EDTA, 5 mM β-mercaptoethanol for 1h and subsequently against b) 50 mM Mes pH 6.2, 0.3 M NaCl, 1 mM EDTA, 5 mM β-mercaptoethanol to avoid protein precipitation. After dialysis, DENV 3 RdRp proteins retaining their N-terminal His_6 tag (hereafter named H_6 DENV 3 RdRp), were diluted to make final NaCl concentration to 50mM and then subjected to cation exchange chromatography, using a Source 15 S column (GE Healthcare, Sweden) with a volume of 20 ml. Tagged proteins were eluted using a linear gradient ranging from 0.15 to 1.5 M NaCl in buffer B (50 mM Mes pH 6.2, 0.05 M NaCl, 5 mM β-mercaptoethanol, 1 mM EDTA). In the case of H_6 DENV 3 RdRp, two distinct fractions (labeled F1 and F2) were eluted from the cation exchange chromatography and were separately purified with gel filtration chromatography (see below). Alternatively, after elution from the HisTrap HP column the H_6 DENV 3 RdRp protein was treated with thrombin overnight (~20-22 h) at 4°C in order to remove its N-terminal hexa-histidine (His_6) tag. The protein fraction was first concentrated and diluted in 20mM Tris HCl pH8.0, 50mM NaCl, 5mM BME (buffer D) to remove imidazole. Using 1U thrombin per 100 µg of protein, proteins was digested in buffer D. After treatment with thrombin, the DENV3 RdRp protein retains residues GSHMLDN at its N-terminus, which are derived from the vector used for cloning. Finally both the tagged and untagged protein were concentrated by ultrafiltration with a molecular weight cutoff of 30 kDa (Milipore, Volketswil, Switzerland), a final gel filtration chromatography step (HiPrep 16/26, Superdex 200) was carried out in buffer C (20 mM Tris HCl at pH 6.8, 0.25 M NaCl, 1 mM EDTA, 2 mM β-mercaptoethanol and 0.1% w/v CHAPS) and both the tagged and untagged proteins were concentrated to ~11 mg ml^{-1} and used for crystallization assays.

Appendix List A2

Mutagenesis

Surface structural mutagenesis on DENV3 RdRp

With the aim to improve the WT catalytic domain activity and stability (Table 1), the surface bulky residues were investigated for mutagenesis studies and based on the structure of WT DENV 3 RdRp (see chapter 5). The surface bulky residues were suspected to cause protein instability by surface polar residues repulsion. Firstly, one mutant that is W846N has remarkably different characteristic compared to the WT. The W846N mutant improves the thermo-stability and protein solubility (achieves 20 mg/ml in a low salt buffer, Table 1& 2). Secondly, most mutants such as R352E/R357E were contributed only to protein expression yield. Thirdly, none of these mutants contribute to the improvement of enzymatic activity compared to those of FL DENV2 NS5 protein. In general, good soluble protein with high protein concentration and in low salt buffer could be more tolerate to PEG crystallization screening conditions. Therefore, W846N mutant offers screening new crystal growth conditions for producing new crystal packing systems, which possibly help to improve the current crystal model (missing residues) or shorten duration of crystal growth. Furthermore, good solubility of W846N mutant is also useful for biophysical studies: (1) 2D NMR-ligand screening or (2) Isothermal Titration Calorimetry (ITC) for ligand binding assay. These approaches often require protein in high concentration in low salt concentration while still remains soluble and stable in the solution.

Rational mutant design on WNV FLNS5

Mutational analyses were also performed on WNV FLNS5, after initial crystallization screenings have unmet crystal growth condition for the WT protein. The mutant designs were based on the limited proteolysis on DENV2 FLNS5 (R264), W846N DENV3 RdRp solubility improvement and surface entropy reduction (table 3). The aim of the mutational analysis was: (1) to produce proteolytic resistant FL (cleavage site at R264) and improve protein solubility in low salt buffer (W846N). Mutant R264E unable to prevents proteolysis whereas mutant W846N slightly improves protein solubility in low salt buffer (4.5 mg/ml compared to WT-2.9mg/ml) (Table 1) and (2) to produce lower conformational entropy protein based on the surface entropy reduction (SER) prediction. Mutations were predicted based on algorithm incorporating a conformational entropy profile, secondary structure and sequence conservation. The method has been successfully used to crystallize more than 15 novel proteins. We designed 8 constructs and 1 was selected for crystallization screening, based on the purity and score of SER but no crystal growth condition was encountered yet (Table 3, no 5).

Table 1. Surface structural mutagenesis of DEN3 RdRp

No	Mutant	Protein	Protein yield estimation	Comment	Enzymatic activity (SPA based)
1	WT DEN3RdRp	DEN3 RdRp	3-5mg/L	achieve 10 mg/ml in 0.25M NaCl	20 fold < FL NS5
2	R352E/R357E DEN3RdRp	DEN3 RdRp	7-8mg/L	-	Same as WT
3	R357E DEN3RdRp	DEN3 RdRp	2-3mg/L	-	Same as WT
4	L880N DEN3RdRp	DEN3 RdRp	2-3mg/L	-	Same as WT
5	W823N DEN3RdRp	DEN3 RdRp	0.5mg/L	-	Same as WT
6	W846N DEN3RdRp	DEN3 RdRp	2-4mg/L	achieve 20 mg/ml in 0.15M NaCl	Same as WT
7	P864D DEN3RdRp	DEN3 RdRp	1-2mg/L	-	Same as WT
8	I524T/P525D DEN3RdRp	DEN3 RdRp	0.3-0.4mg/L	-	Same as WT
9	W846N/L880N DEN3RdRp	DENV 3 RdRp	1-2mg/L	-	Same as WT
10	W846N/W864N /L880N DEN3RdRp	DEN3 RdRp	3-4mg/L	-	Same as WT

Table 2. Termostability test by thermofluor (Cummings et al., 2006)

Protein	**Melting temperature**
W846N DENV3RdRp	44°C
W846N/L880N DENV3RdRp	44°C
L880N DENV3 RdRp	42°C
DEN1 RdRp & DENV3 RdRp	41°C
DENV2 FLNS5	40°C

Procedure:

Use 5 μg protein in 25 μl buffer 50 mM Tris HCl pH 7.5, 150 mM NaCl, 10% G, 2 mM BME for each well. Add 1 μl sypro orange (Invitrogen) to 40 μl protein buffer (40x dilution). Vortex the solution. Then add 1 μl from the 40 μl stock into 25 μl for each well to make 1:1000 dilutions. Spin down for 1 min 6000 x g and remove solution to a new tube and pipette 25 μl into each well. Briefly spin the plate (short spin to ensure all solution spin down to the well). Start reading using Bio-Rad5iQ. More than 1°C difference is considered significant melting temperature change.

Table 3. Surface entropy reduction on WNV FLNS5 (Goldschmidt et al., 2007)

no	WNV FLNS5 mutant	Protein expression yield	Design rational	Remark	SER core, based on WT
1	262264 EKK-->AAA		SER prediction		6.93
2	898-900 EEK-->AAA		SER prediction		5.69
3	898-900 EEK-->AEA		SER prediction		5.69
4	898-900 EEK-->AEK		SER prediction		5.69
5	478-480 EKK-->AAA		SER prediction	Selected for crystallization screening	5.16
6	408-409 KK-->KA		SER prediction		4.98
7	433-435 EEQ-->AAQ		SER prediction		4.72
8	433-435 EEQ-->EAQ		SER prediction		4.72
9	R264E	~WT	Limited proteolysis on DEN		N.A
10	W846N	~WT	DEN3 RdRp solubility		N.A

Reference List

Cummings,M.D., Farnum,M.A., and Nelen,M.I. (2006). Universal screening methods and applications of ThermoFluor (R). Journal of Biomolecular Screening *11*, 854-863.

Goldschmidt,L., Cooper,D.R., Derewenda,Z.S., and Eisenberg,D. (2007). Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. Protein Science *16*, 1569-1576.

Appendix List A3

Compound soaking list (for internal use only)

Table 1. List of compounds and parameters that wereused in this thesis (chapter 4).

No	Actual Compound ID	Reference ID used in this soaking study	M.W	Soaking concentration	Soaking time	Data collection
1		72	189.17	4 mM	24 h, 48 h, 4°C	Yes
2		43	235.29	4 mM	24 h, 48 h, 4°C	Yes
3		57	327.71	4 mM	24 h, 48 h, 6 d, 4°C	Yes
4		45	245.69	4 mM	24 h, 48 h, 6 d, 4°C	Yes
5		48	301.22	4 mM	24 h, 48 h, 4°C	Yes
6		47	223.25	4 mM	24 h, 48 h, 4°C	
7		30	201.22	4 mM	24 h, 48 h, 4°C	Yes
8		32	239.25	4 mM	24 h, 48 h, 6 d, 4°C	Yes
9		63	256.69	4 mM	24 h, 48 h, 4°C	
10		11	308.32	4 mM	24 h, 48 h, 4°C	
11		25	308.36	4 mM	24 h, 48 h, 6 d, 4°C	Poor diffraction
12		49	225.27	4 mM	24 h, 48 h, 4°C	Yes
13		150	305.34	4 mM	24 h, 48 h, 4°C	Yes
14		125	232.34	4 mM	24 h, 48 h, 6 d, 4°C	Yes
15		149	210.23	4 mM	24 h, 48 h, 6 d, 4°C	Poor diffraction
16		71	216.2	4 mM	24 h, 48 h, 4°C	
17		120	220.18	4 mM	24 h, 48 h, 4°C	Poor diffraction
18		135	169.14	4 mM	24 h, 48 h, 6 d,4°C	Yes
19		84	251.14	4 mM	24 h, 48 h, 6 d,4°C	Yes
20		132	451.5	1-5 mM	18-20 h, 4°C	Yes
21		519	459.53	1-5 mM	18-20 h, 4°C	Yes
22		555	656.7	1-5 mM	18-20 h, 4°C	Yes
23		671	565.7	1-5 mM	2h, 22 h, 4°C	Yes

Appendix List A4

Potential RNA sequence for DEN RdRp + template interaction

Novel mechanism for (-) strand RNA synthesis was proposed by Filomatori et al: viral polymerase binds SLA at the 5'end of the genome and reaches the site of initiation at the 3'end via long range RNA-RNA interactions. With these finding, the following RNA templates were proposed for the future follow up and experiment design to obtain enzyme-template-substrate ternary complex. The template tunnel of DEN polymerase could accommodate 5-7 mers of RNA template, based on the FMDV and Φ6 enzyme-template-substrate ternary complexes. Two approaches can be adopted: a) use either SLA of the 5'UTR or 3'UTR for screening co-crystallization condition with the following template length (see below, i) and b) use SLA to identify the optimal binding region and length for crystallization assay. The SLA is ~70 mers and template tunnel of dengue polymerase can fit 5-7 mers template. Therefore, experiment can be designed for this objective (see below, ii).

i. RNA templates design for Co-crystallization assays

SLA 5'--3'

1. AGUUGUUA
2. AGUUGUU
3. AGUUGU
4. AGUUG

3' UTR 5'--3'

1. CAGGUUCU
2. AGGUUCU
3. GGUUCU
4. GUUCU

ii. Optimal binding region identification of SLA

a. First 70 nucleotides of SLA are divided into 7 parts:

1. AGUUGUUAGU
2. CUACGUGGAC

3. CGACAAAGAC
4. AGAUUCUUUG
5. AGGGAGCUAA
6. GCUCAACGUA
7. GUUCUAACAG

b. Overlap nucleotides between the first 70 nucleotides-

1. UAGUCUAC
2. GGACCGAC
3. AGACAGAU
4. UUUGAGGG
5. CUAAGCUC
6. CGUAGUUC