CRYSTAL STRUCTURE OF SINDBIS VIRUS PRECURSOR FORM OF NONSTRUCTURAL PROTEIN 2/3

by

GYE HWA SHIN

A thesis submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

And

The Graduate School of Biomedical sciences

University of Medicine and Dentistry of New Jersey

In partial fulfillment of the requirements

For the degree of

Master of Science

Graduate Program in Biochemistry

Written under the direction of

Dr. Joseph Marcotrigiano

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

[Oct, 2010]

ABSTRACT OF THE THESIS

CRYSTAL STRUCTURE OF SINDBIS VIRUS PRECURSOR FORM OF NONSTRUCTURAL PROTEIN 2/3

By GYE HWA SHIN

Thesis Director:
Dr. Joseph Marcotrigiano

Alphaviruses are a globally distributed group of enveloped, positive-sense RNA animal viruses. The 5' two-thirds of the RNA genome is translated into a single, large polypeptide of ~2500 amino acids. This polypeptide is autoproteolytically cleaved into four subunits, termed nonstructural proteins (nsP) 1-4, which form the replication complex. The nsP1 is the enzyme responsible for capping the viral mRNA and a membrane associated protein. The nsP2 has RNA helicase, triphosphatase activities and is an autoprotease responsible for cleaving the nonstructural polyprotein. The nsP4 is an RNA-dependent, RNA polymerase. The function of nsP3 remains unknown, however, it is the only viral protein phosphorylated by unknown cellular kinases. The nsP3 contains a macro or X domain that binds ADP-ribose and is highly conserved with homologues in rubella virus, hepatitis E virus and coronaviruses, including severe acute respiratory syndrome (SARS) virus. The protease domain of viral nsP2 protein cleaves the nsP1234 polyprotein replication complex into functional proteins. During replication of alphavirus genome, the positive strand of RNA genome is transcribed into a minus strand by the cleavage intermediate nsP123 and nsP4. Cleavage between nsP2/3 causes a template

switch from minus strand to positive strand. Although the nonstructural protein 2/3 is essential for RNA replication, its function in viral replication is still not well understood. Here we report the crystal structure of Sindbis virus precursor form of nsP2/3 at 2.94Å resolution. The structure of Sindbis nsP2/3 protein consists of four distinct domains, which are the protease domain, methyltransferase like domain, macro domain and zinc binding domain. Surprisingly, nsP3 contains a novel folding with a previously uncharacterized zinc coordination site. In addition, the cleavage site between nsP2 and nsP3 is buried the two proteins and not accessible for proteolysis. Furthermore, the loop connecting the macro and zinc-binding domains makes direct contact with and causes a structure shift in the nsp2 protease domain. We have observed considerable differences between the pre-cleavage and post-cleavage forms of nsp2/3. The structure suggests that a conformational change in nsP2/3 is required in order for the nsP2 protease to gain access to the nsP2/3 cleavage site.

Acknowledgement

Firstly, I would like to thank my family, husband, mother, father and my daughter being there for me and supporting me all along. My family has always encouraged and helped me during my studies.

Secondly, I would like to thank my advisor, Dr. Joe Marcotrigiano, for his guidance and support throughout my research. During my master degree in Dr. Joe's lab, I have learned a lot of new techniques, improved skills, and explored a new field of study called X-ray crystallography. I would also like to thank Dr. Abram Gabriel and Dr. Vikas Nanda for their advice, suggestions and comments on my thesis. As a graduate director, Dr. Gabriel gave me a good guidance and encouragement to study in biochemistry department.

I specially thank Dr. Matthew Miller and my lab mates, Jill and Fuguo for their invaluable help and support.

Table of Contents

Abstract of Thesis ii

Acknowledgements and Dedication iv

Table of Contents v

List of Tables vi

List of Figures vii

Introduction 1

Materials & Methods 7

Results 11

Discussion 37

References 40

Lists of Tables

Table 1. X-ray crystallographic Data collection and refinement statistics 12

List of Illustrations

Figure 1. Replication of alphavirus genome 2

Figure 2. Schematic representation of various viral and cellular X domain containing proteins, highlighting conserved motifs 4

Figure 3. Genome organization of alphaviruses 6

Figure 4A. Purification of GST-nsP2/3 8

Figure 4B. SDS-PAGE analysis of GST-nsP2/3 protein purification 8

Figure 5. Crystallization of Sindbis virus nsP2/3 11

Figure 6A. Ribbon diagram of Sindbis virus nsP2/3 colored from blue (N terminus) to red (C terminus) 13

Figure 6B. Superposition of representations of Sindbis virus nsP2/3 with VEEV nsP2 protein 15

Figure 7. Electrostatic surface of the Sindbis nsP2/3 17

Figure 8. Structure of Sindbis nsP2 protease domain 18

Figure 9. Structure of Sindbis nsP2 methyltransferase like domain 21

Figure 10. Structure of Sindbis virus nsP3 macro domain 24

Figure 11. Structure of Sindbis virus nsP3 zinc binding domain 28

Figure 12. Location of temperature Sensitive Mutants 31

Figure 13. Sequence alignment of Alphaviruses 34

Chapter 1

Introduction

The alphavirus is a member of the *Togaviridae* family, a group of positive strand RNA viruses that is an important human and animal pathogen with global distribution. Some alphaviruses can cause encephalitis in animals and humans, while others cause mild symptoms of disease such as fever, rash and arthritis (Laine et al., 2004).

Sindbis virus (SIN) is a prototype of alphaviruses and has been intensively studied for the replication of alphaviruses (Lemm and Rice, 1993; Lemm et al., 1994; Shirako and Strauss, 1994; Gorchakov et al., 2008). Sindbis virus was first isolated in August 1952, from a pool of mosquitoes, and was transmitted to vertebrate hosts by mosquitoes (Taylor 1955). The Sindbis virus genome is a single, positive-stranded RNA of 11703 nucleotides. It includes a methylguanylate cap at the 5' terminus and a poly-adenylated tail at the 3' terminus. The genomic 49S RNA produces mRNA in infected cells. After a cell is infected, the viral nonstructural proteins (nsPs) nsP1, nsP2, nsP3 and nsP4 are produced by translation of the genomic RNA (Strauss et al., 1994). Initially, the nonstructural proteins are synthesized as two polyproteins, P123 and P1234. Alphaviruses terminate at an opal codon between nsP3 and nsP4 to produces the P123 polyprotein including nsP1, nsP2 and nsP3. Polyprotein P1234 is randomly produced upon read through of the opal codon and contains all four nsP sequences. Sequential cleavage event of the nonstructural polyproteins was known to be an important process for regulation of minus and plus strand synthesis. During replication of the alphavirus genome, the positive strand RNA genome is transcribed into a minus strand by the cleavage intermediate nsP123 and nsP4 (Strauss et al., 1994). Cleavage between nsP2/3 causes a template switch from minus strand to

positive strand synthesis. The structural proteins are translated from a 26s subgenomic mRNA which has the same sequence as the 3' end one third of genomic RNA. The synthesis of the negative and positive strand and transcription of the subgenomic RNA is a highly regulated process (Fig. 1).

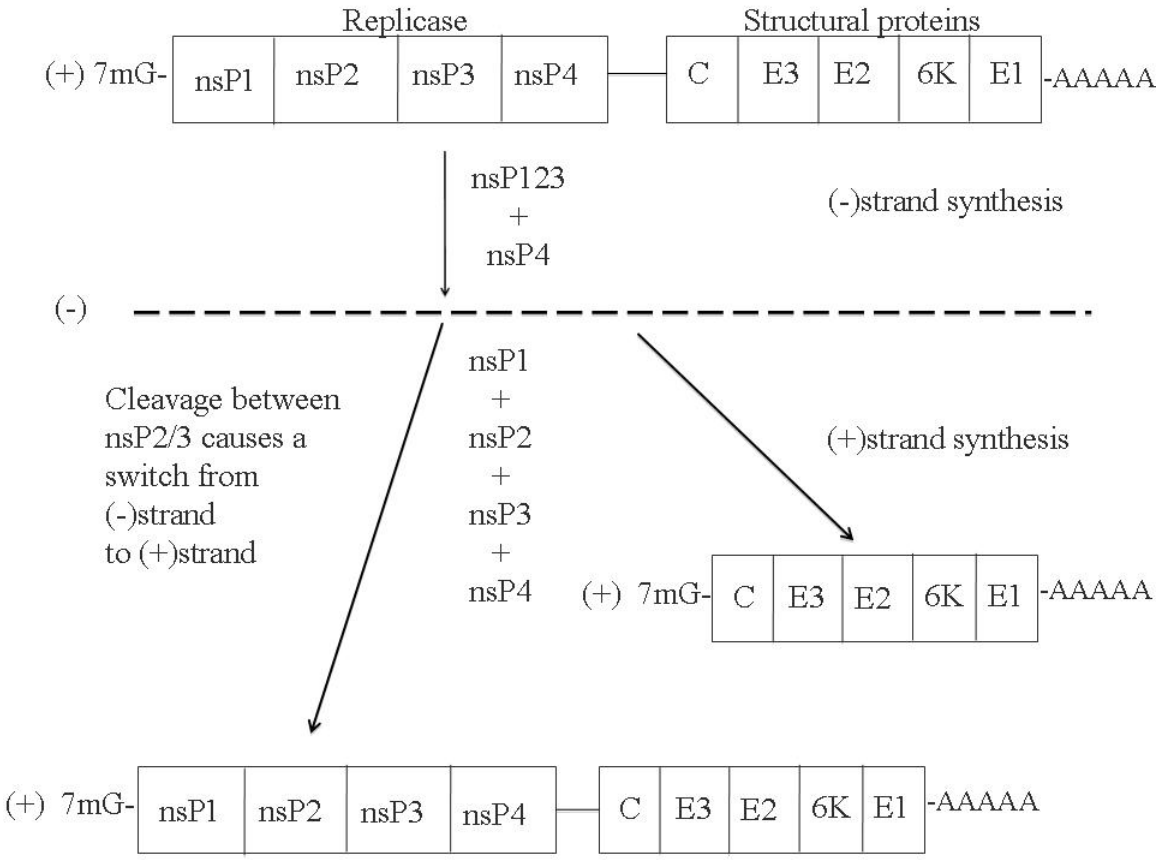


Figure 1. Replication of alphavirus genome. The positive strand, RNA genome is transcribed into a minus strand by the cleavage intermediate nsP23 and nsP4. The fully proteolyzed nonstructural protein (nsP1, nsP2, nsP3, nsP4) synthesize progeny, positive strands and subgenomic RNA, which encode for the viral structural protein.

Although the viral protein processing has been extensively studied, the molecular mechanism of the nsP2/3 site cleavage and sequential processing are not well known.

The nsP1 is a membrane-associated protein and has methyltransferase (MTase) and guanylyltransferase activity. It is involved in capping the viral RNA. Initiation of minus strand synthesis occurs by a subsequent interaction between nsP1 and nsP4 protein (Ahola and Kaarianinen, 1995; Wang et al., 1996). It is also known that nsP1 is associated with the plasma membrane, endocytic vesicles, and lysosomal vesicles (Peränen et al., 1995). The nsP2 is composed of three domains and has multiple functions. The amino terminal half is an RNA helicase, RNA dependent 5' triphosphatase and nucleoside triphosphatase (Gomez et al., 1999; Vasiljeva et al., 2000). The carboxy terminal portion of nsP2 contains a papain-like cysteine protease domain and a methyltransferase-like domain of unknown function. The C-terminal of nsP2 contains a papain-like protease domain for autocatalytic processing of P1234 and P123 polyproteins and can sequentially cleave the nonstructural polyproteins into each nonstructural protein (nsP1, nsP2, nsP3 and nsP4). It is also involved in regulating synthesis of the subgenomic RNA and 50% of the translated nsP2 is translocated into nucleus where it shuts off host cell gene expression (Peränen et al., 1990). The nsP4 protein is the viral RNA-dependent RNA polymerase. It also has terminal adenyl transferase activity that is involved in adding the poly (A) tail at the 3' end of the genome (Lemm et al., 1994; Tomar et al., 2006). The nsP3 protein is not well understood in terms of the function and specific roles. There is no precise evidence describing the nsP3 protein function or activity yet, although it is known that nsP3 play an important role in the transcription process at the early stage of the infection. The N-terminal two thirds of nsP3 are well conserved. The amino terminal region contains a macro or X domain (Koonin et

al., 1992). This domain derives its name from a variant of histone H2A, called macroH2A (Pehrson and Fried, 1992). Macro domains are widely distributed throughout all kingdoms of life, including eukaryotic organisms, pathogenic bacteria, such as *Listeria* and *Salmonella* (Karras et al., 2005), archea and single stranded RNA viruses including hepatitis E virus, rubella virus, coronavirus and alphavirus (Fig. 2)

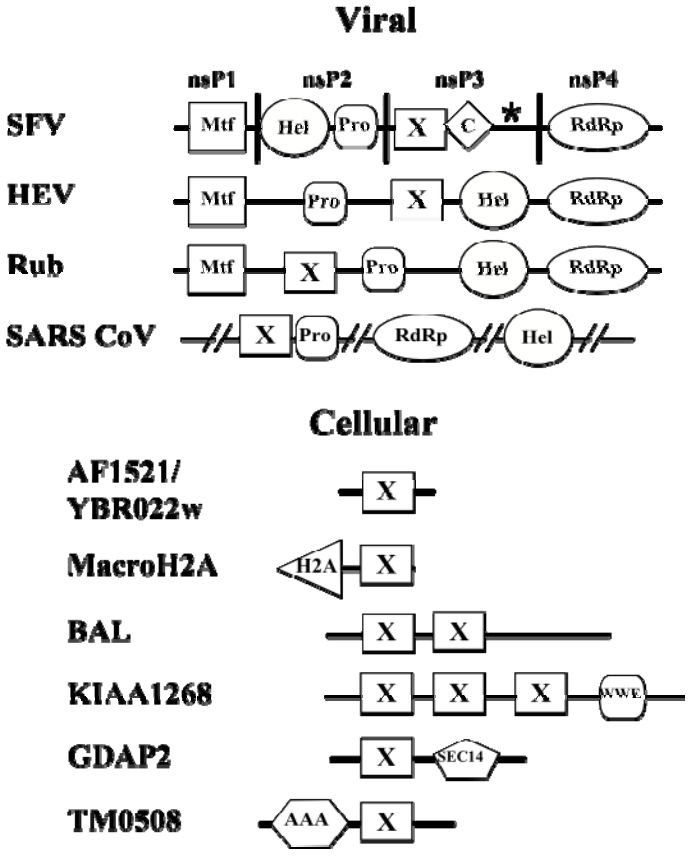


Figure 2. Schematic representation of various viral and cellular X domain containing proteins, highlighting conserved motifs. The four alphavirus nonstructural proteins are marked with a vertical line and labeled nsP1-4. The conserved motifs are as follows: X domain (X), RNA methyltransferase (Mtf), RNA helicase (Hel), papain-like protease (Pro), RNA-dependent, RNA polymerase (RdRp), histone 2A (H2A), a protein interaction module (WWE), domain in homologues of a *Saccharomyces cerevisiae* phosphatidylinositol transfer protein (SEC14), AAA ATPase (AAA), conserved domain found only in alphaviruses (C), location of several nsP3 phosphorylation sites (*).

The macro domain was initially characterized to contain enzymatic activity of ADP-ribose-1”-phosphate phosphatase. Later it was shown that most domains lack the phosphatase activity but retain ADP-ribose binding site. The ADP-ribose plays an important role for posttranslational modification in DNA repair, transcriptional activation, chromatin modification and long term memory (Karras et al., 2005). However, ADP-ribose binding affinity is not quite consistent in different organisms. For example, ADP-ribose-1”-phosphate phosphatase high activity shows in Af1521 (*A. fulgidus*) and SARS-CoV, but low activity in Semliki Forest virus (SFV) alphavirus. The macro domain of SFV could bind to poly ADP ribose (PAR) as well. PAR proteins are known to be involved in the cellular response to viral infection, inflammation and stress as well as cellular processes involving DNA repair and programmed cell death. The carboxy terminal region is not conserved well among alphaviruses and is phosphorylated on serine and threonine sites.

Here we report the crystal structure of the Sindbis virus precursor form of nsP2/3 at 2.94 Å resolution. The structure of Sindbis nsP2/3 protein consists of four distinct domains, which are the protease domain, methyltransferase like domain, macro domain and a previously unknown and uncharacterized zinc binding domain (Fig. 3).

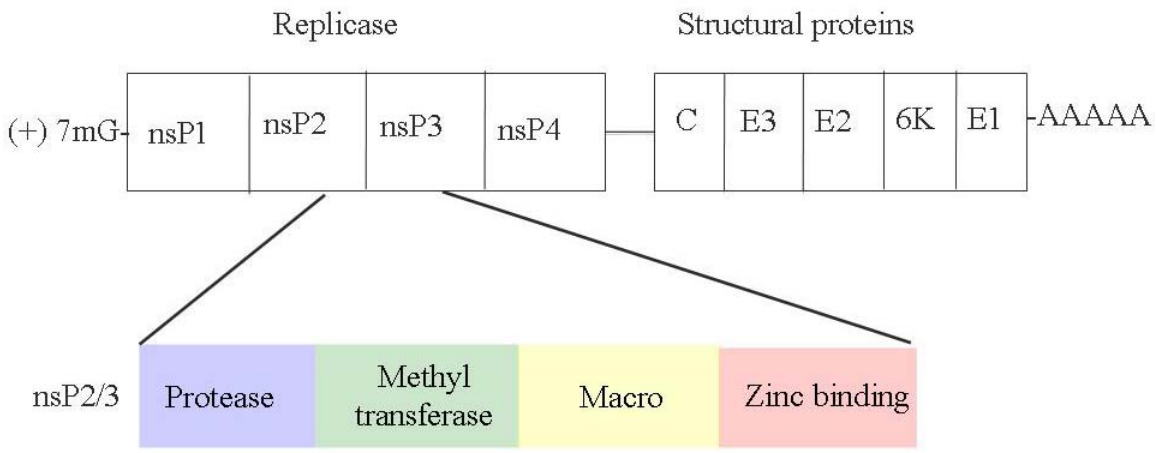


Figure 3. Genome organization of alphaviruses. The highlighted domains indicate Sindbis nonstructural protein 2/3 (nsP2/3).

The zinc binding domain is a novel fold. In addition, the cleavage site between nsP2 and nsP3 is buried between the two proteins and not readily accessible for proteolysis. Furthermore, the loop connecting the macro and zinc-binding domains makes direct contact with and causes a structure shift in the nsP2 protease domain. We have observed considerable novel aspects within the pre-cleavage forms of nsp2/3. The structure suggests that a conformational change in nsP2/3 is required in order for the nsP2 protease to gain access to the nsP2/3 cleavage site.

Chapter 2

Materials and Methods

2.1 Cloning, Expression and Purification of Sindbis nsP2/3

DNA construct was designed as wild type, 470 residues deletion from the N-terminus of Sindbis nsP2 and 13 amino acid deletion from residue number 341 of nsP3 (wt dN470dC13). DNA coding for nsP2/3 (wt dN470dC13) was amplified by polymerase chain reaction (PCR) and cloned into pGEX-6p-1 vector (GE lifesciences). *E. coli* strain UT5600 were transformed with pGEX-6p-1 vector, grown at 30°C in Lysogeny broth (LB, 10 g Bactotryptone, 5 g Yeast extract, 10 g NaCl) medium overnight. The next day the overnight culture was diluted 1:100 into one liter of LB plus antibiotics grown to a density of 0.7~0.8 OD at 600 nm, chilled at 4°C for 45 minutes, and induced by 0.2 mM isopropyl β-D-1 thiogalactopyranoside (IPTG) at 18°C overnight. Cells were harvested by centrifugation at 4500 rpm. Cell pellets were resuspended in 50 mM HEPES, 150 mM KCl and 10% glycerol and lysated by homogenization in french press (Avetsin) and the supernatant were centrifuged at 16000 rpm. The proteins were purified using glutathione S-transferase (GST) affinity column (GE healthcare) (Fig. 4A). Fig. 4B shows SDS-PAGE analysis of GST-nsP2/3 protein purification.

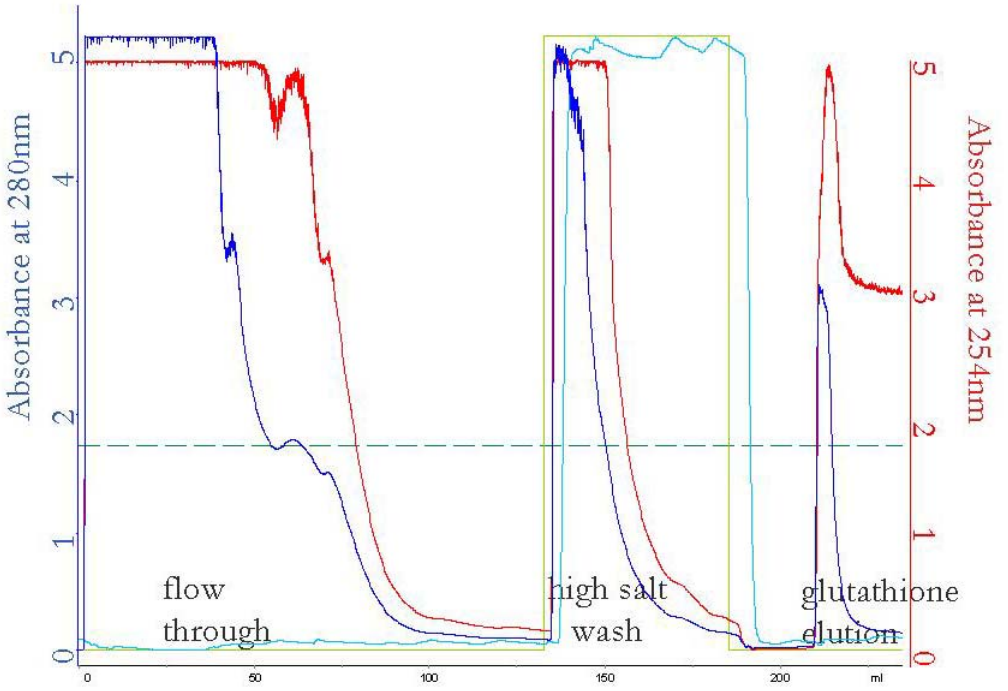


Figure 4A. Purification of GST-nsP2/3. Chromatogram showing the absorbance at 280nm (blue) and 254nm (red) as a function of time (x axis).

Figure 4B. SDS-PAGE analysis of GST-nsP2/3 protein purification.

Purification sequentially required use of a hydroxyapatite column (Biorad), desalting column (GE healthcare) and heparin column (GE healthcare). The first purification was performed on a 5 ml GST prep column. The proteins were eluted with 0.1 M Tris-HCl (pH8), 150 mM KCl, 15 mM glutathione and 5% glycerol. After GST purification, overnight dialysis was performed against 20 mM HEPES (pH7.5), 150 mM KCl, 5% glycerol and 2mM dithiothreitol (DTT). The next day, the dialyzed sample was centrifuged at 16000 rpm for 15 min to remove precipitant. After centrifugation, the protein was purified by hydroxyapatite column. A linear gradient was used to elute the protein using loading buffer A (5 mM phosphate buffer (pH7.5), 150 mM KCl, 5% glycerol) and elution buffer B (200 mM phosphate buffer (pH7.5), 150 mM KCl, and 5% glycerol). Normally, the protein eluted at 70 mM~100 mM phosphate buffer. Fractions were analyzed by SDS-PAGE with Coomassie blue staining and the best fractions were pooled. The protein was desalted using a Hiprep 26/10 desalting column (GE healthcare). The composition of the desalting buffer was 50 mM HEPES (pH7.5), 50 mM KCl and 5% glycerol. Finally, the protein was purified by heparin column using 50 mM HEPES (pH7.5), 1 M KCl and 5% glycerol. Before concentration of the protein, it was desalted by using 20 mM HEPES, 150 mM KCl, 5% glycerol and 1 mM Tris 2 carboxyethyl phosphine (TCEP). The protein was then concentrated to 15 mg/ml. The concentration of protein was determined by Bio-Rad assay.

2.2 Crystallization and crystal freezing

Crystallization screens were done using Hampton index kit. After optimization, the best crystallization was found in 2 M ammonium sulfate and 0.1 M 2-(N-morpholino)

ethane sulfonic acid (MES) (pH6.5). Crystals were soaked in a cryoprotectant solution containing 2 M ammonium sulfate, 0.1 M MES (pH6.5), 35% xylitol and 5% Trimethylamine (TMAO) for 30 sec ~ 1 min before being flash cooled in liquid nitrogen.

2.3 Data collection and structure determination

Diffraction data were collected at beamlines X25 and X29 at Brookhaven national lab synchrotron light source. The crystal diffracted to a resolution of 2.94 Å. Integration of the different data sets, scaling and merging of the intensities were performed using HKL2000 program. Determination of the phasing was accomplished by a combination of molecular replacement and single wavelength anomalous dispersion of zinc atom. The molecular replacement was done with Phenix program. The map was improved with several rounds of manual model building using Coot and Phenix program.

Chapter 3

Results

3.1 Protein purification and crystallization

The cloned insert contains 662 amino acids from Sindbis nsP2/3. It is a monomeric, which was confirmed by using a gel filtration column (Superdex 200) that estimated a molecular weight of 78 kD. The final nsP2/3 protein construct was determined by analyzing the results of limited proteolysis experiment using trypsin digestion (wt dN470dC13), with digestion products verified by mass spectrometry. This construct was expressed in *E. coli*. The protein was purified using four different columns and crystallized (Fig. 5).

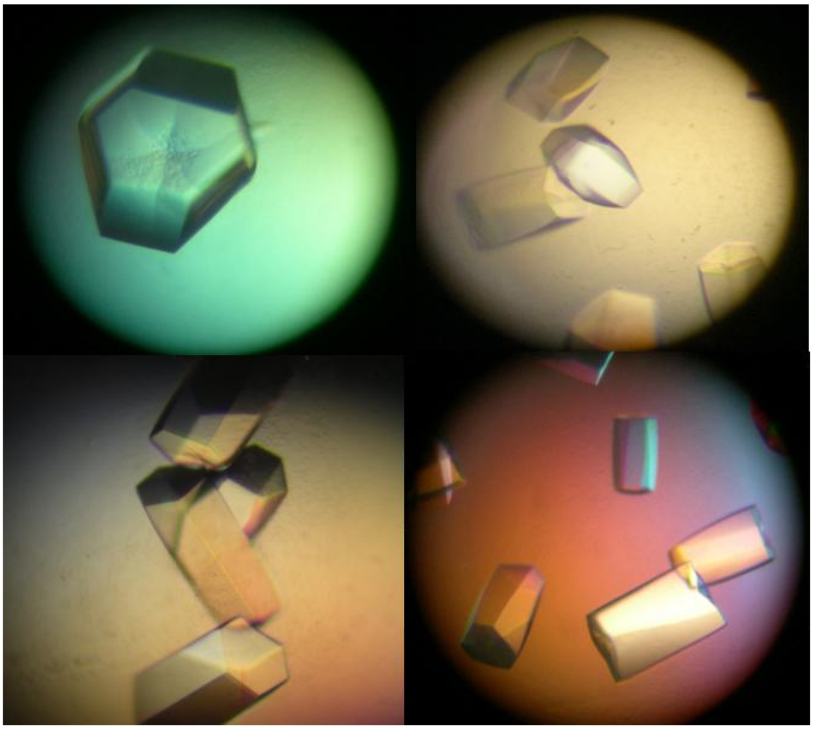


Figure 5. Crystallization of Sindbis virus nsP2/3. Crystals of wild type nsP2/3 construct (wt dN470dC13) in 2M $(NH4)_2SO_4$, 0.1M MES(pH6.5).

Analysis of the diffraction data indicated that the Sindbis nsP2/3 crystal belonged to the $P3_121$ space group. The asymmetric unit contains three molecules of nsP2/3, and crystal solvent content was determined to be ~75% based on the Mattews coefficient (Vm = 4.8 $Å^3$/Da) (Mattews, 1968). Crystal is diffracted to 2.94 Å resolution. Diffraction data were collected at beamline X-25 at Brookhaven national lab synchrotron light source. The nsP2/3 structure was solved by a combination of SAD and molecular replacement. The final model was refined to an R_{factor} and R_{free} of 21.5% and 25.2%, respectively. Structure refinement statistics are summarized in Table 1.

Table 1. X-ray crystallographic Data collection and refinement statistics

Summary of data collection and refinement statistics.

Data Collection						
Wavelength (Å)	Resolution (Å)	Reflections measured/unique	Completeness (%)	R_{sym} [a] (%)	I/σ(I)	Phasing Power[b] (anomalous)
λ= 1.28	50-2.94	356598/96912	99.7(99.5)	0.074 (0.629)	18.29(2.02)	2.34

Refinement					
Resolution (Å)	Cut-off	Reflections	Completeness (%)	R_{cryst} [c] (%)	R_{free} [d] (%)
48.26-2.94	|F|/σ|F|>2	90936	93.57	21.5	25.2

root mean square deviations			
Bond lengths	Bond angles	Thermal parameters mainchain atoms	Thermal parameters sidechain atoms
0.003Å	0.873°	59.779 $Å^2$	61.064 $Å^2$

[a] $R_{sym} = \Sigma|I-\langle I\rangle|/\Sigma I$, where I is observed intensity and $\langle I\rangle$ is average intensity obtained from multiple observations of symmetry-related reflections.
[b] Phasing power = rms ($|F_H|/E$), where $|F_H|$ = heavy atom structure factor amplitude and E = residual lack of closure.
[c] $R_{cryst} = \Sigma||F_{obs}|-|F_{calc}||/\Sigma|F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively.
[d] R_{free} is the same as R_{cryst}, but is calculated with 10% of the data.

3.2 Overall structure of Sindbis nsP2/3

The polypeptide chain folds into four distinct compact domains (Fig. 6A).

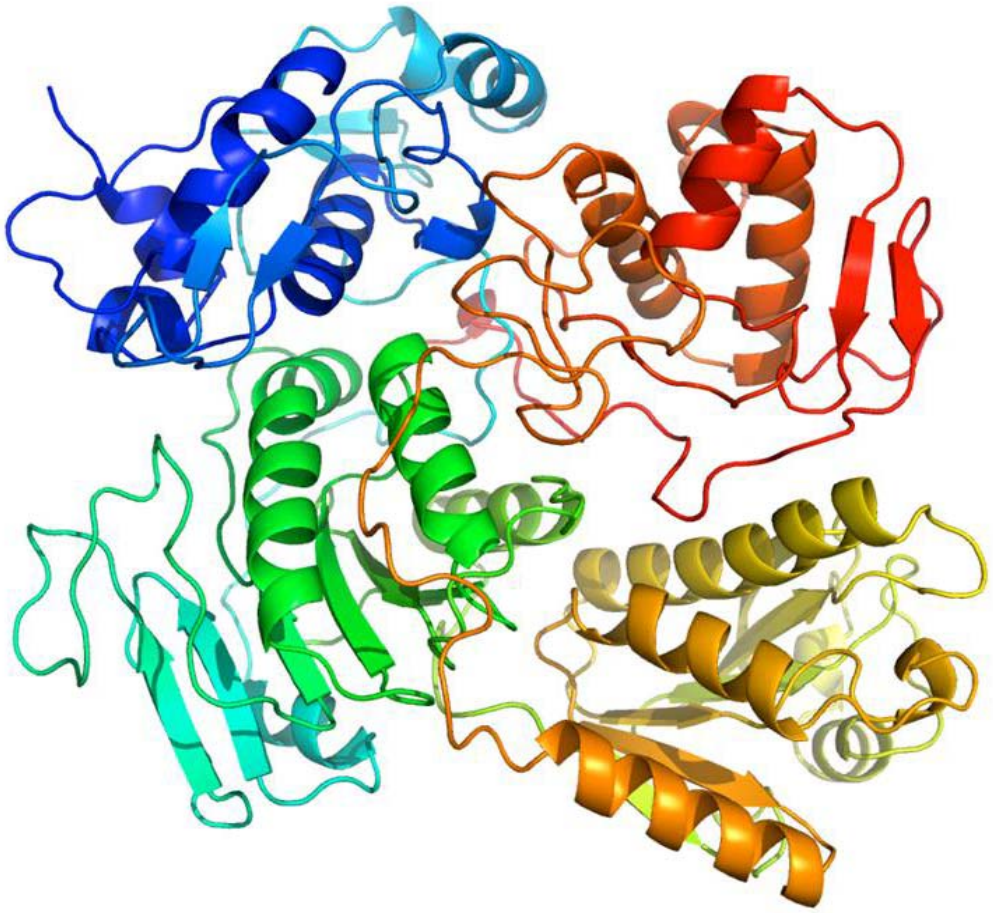


Figure 6A. Ribbon diagram of Sindbis virus nsP2/3 colored from blue (N terminus) to red (C terminus). Sindbis nsP2/3 protein is composed of four different domains: Protease domain, Methyl transferase like domain, Macro domain and Zinc binding domain.

The N-terminal protease domain of Sindbis nsP2/3 is from residue Ala 471(Ala1011) to residue Asn614(Asn1151). The conserved cysteine protease domain contains residue Cys481 and residue His558 which form the catalytic dyad. The protease domain of the Sindbis nsP2/3 protein consists of seven α helices and two regions of short β hairpins structure (Fig. 8). The next domain is the methyltransferase like domain (MT like domain) that extends from Arg615(Arg1155) to Ala807(Ala1347). This domain contains a central twisted six stranded β sheets surrounded by four α helices (Fig. 9). The structure of this domain is similar to the S-adenosyl-L-methionine (SAM) dependant methyltarnsferases superfamily by SCOP (Jennifer and Fiona, 2002). The first domain of Sindbis nsP3 is the macro domain which extends from residue Ala 808 (Ala1348) to residue Ala966 (Ala1506). This macro domain consists of a central twisted six-stranded β sheets surrounded by four α helices on one side and two short helices on the other (Fig. 10). The macro domain has a deep hydrophobic pocket for binding ADP ribose or PAR. Electrostatic surface analysis of the Sindbis nsP3 macro domain shows a highly positively charged patch. Lastly, the second domain of Sindbis nsP3 domain is a novel folded zinc-binding domain, which starts at residue Glu967 (Glu1507) to end of sequence at residue Tyr1132 (Tyr1672). Four cysteine residues from the loop coordinate a zinc ion within the domain (Fig. 11). The coordination of this zinc binding domain is a totally novel fold and the biological role is not yet known.

The cleavage site between nsP2 and nsP3 is buried between the MT like domain and macro domain. It is not accessible for proteolysis. Furthermore, the loop connecting the macro domain and zinc-binding domains makes direct contact with the α2 helix in the MT like domain, causing a kink in the helix and subsequent downstream structural shifts in

the nsP2 MT like domain. When the Sindbis nsP2/3 structure was superimposed with the VEEV nsP2 structure, the MT like domain was shifted down compared to the structure of Sindbis nsP2/3 due to this contact of the connecting loop (Fig. 6A and Fig. 6B).

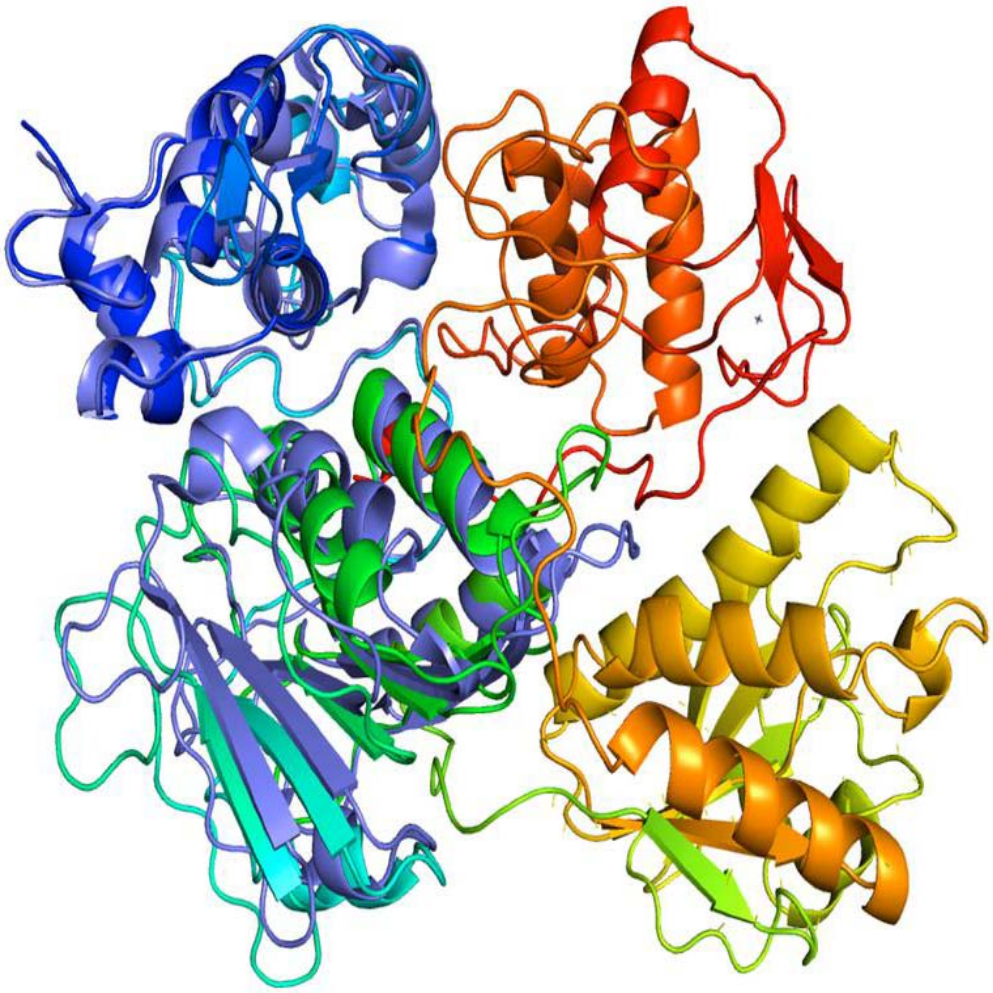


Figure 6B. Superposition of representations of Sindbis virus nsP2/3, the Venezuelan Equine Encephalitis Alphavirus nsP2 protease domain and methyltransferase like domain. The loop connecting the macro domain and zinc binding domain makes direct contact with and causes a shift in the nsP2 methyltransferase-like domain.

Electrostatic surface analysis of the Sindbis nsP2/3 shows that the negatively charged loop connecting the macro domain and zinc-binding domains makes direct contact with the α2 helix in MT like domain, which is half highly positively charged surface and half negatively charged surface. In addition, there is a channel between the MT domain and the macro domain. Interestingly, this channel has a positively charged hole on one side and negatively charged hole on the other, forming a very nice pocket. This channel may have an important function in the viral polyprotein processing (Fig. 7).

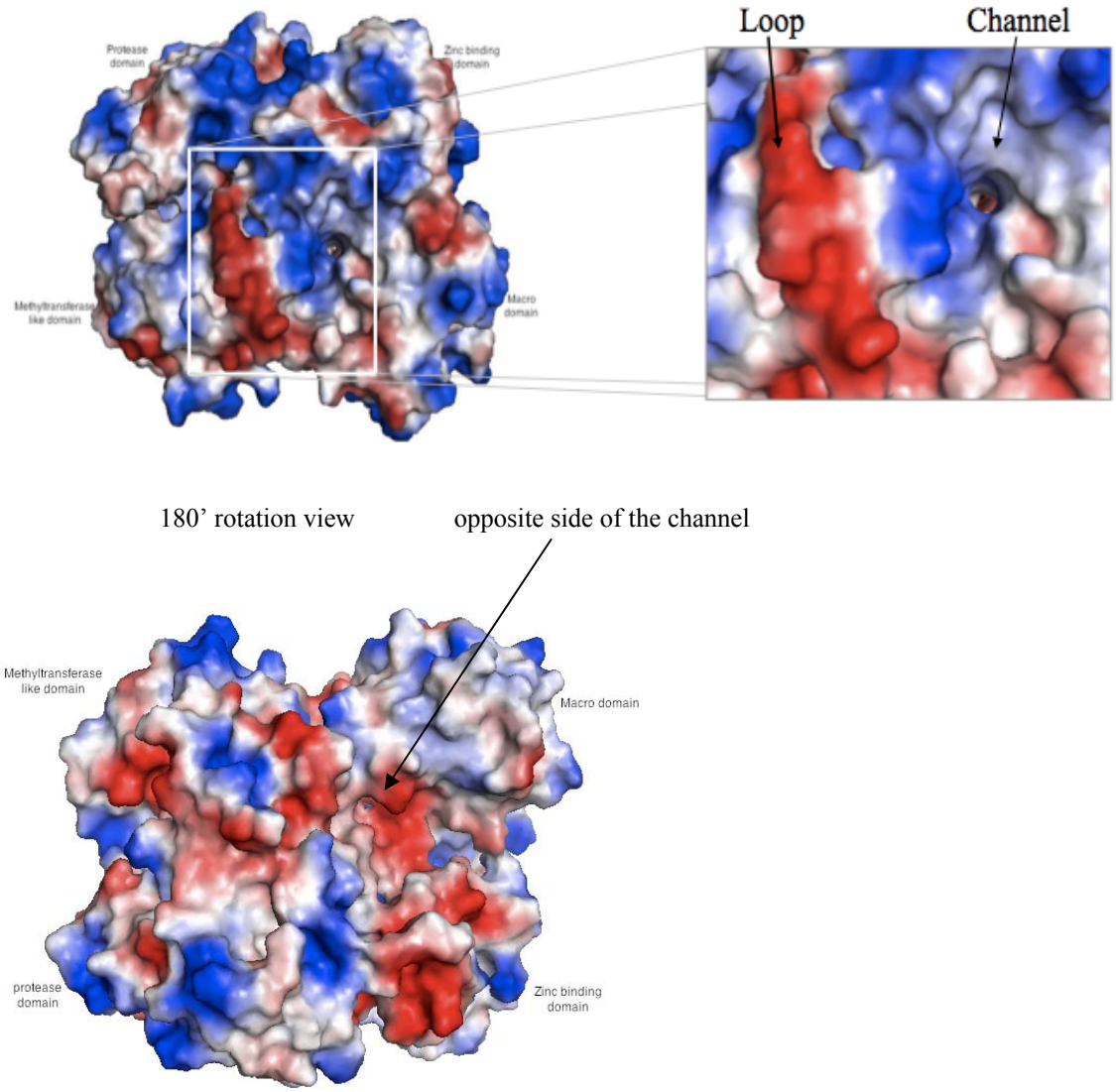


Figure 7. Electrostatic surface of the Sindbis nsP2/3. The loop connects between the macro domain and the Zinc binding domain. The channel is located in macro domain.

3.3 Sindbis nsP2 protease domain

The protease domain consists of helix and random coil regions (Fig. 8).

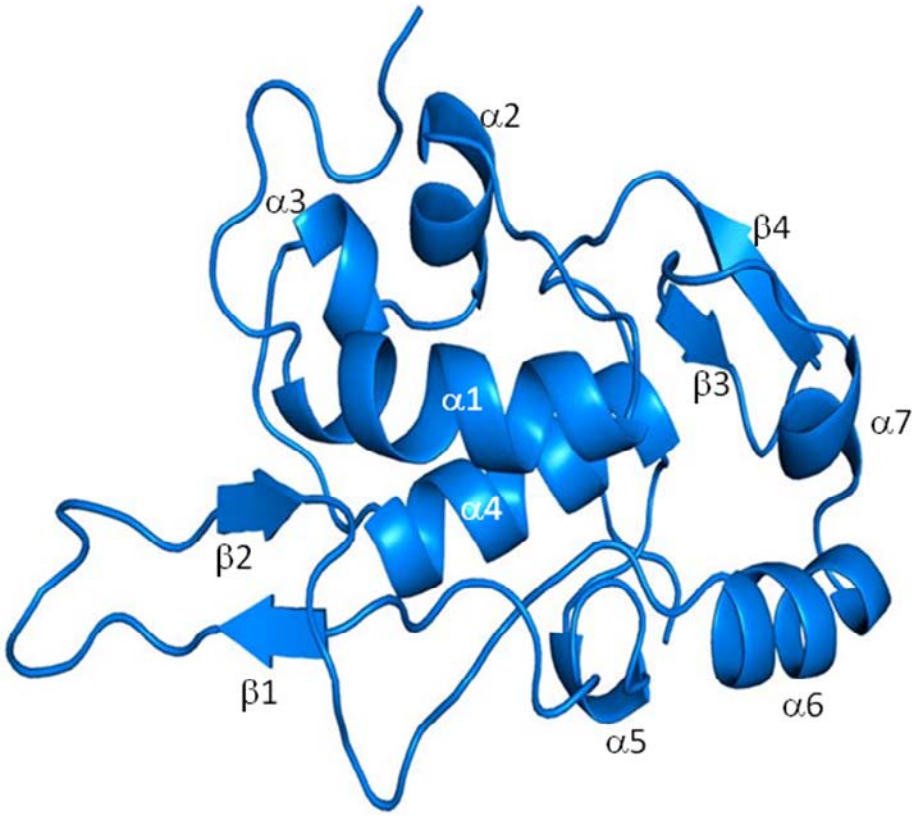


Figure 8. Structure of Sindbis nsP2 protease domain shown in Blue. The protease activity of nsP2 protein is important for cleaving the nonstructural polyprotein (nsP1234) into functional single protein.

Cys481 (VEEV Cys477) and His558 (VEEV His546) form the conserved cysteine protease catalytic dyad within the N-terminal proteolytic domain, which runs from residue Ala471 to residue Asn614. The orientation of the catalytic dyad residues in nsP2 is similar to the catalytic dyad conformation observed in papain (Drenth et al., 1968). The catalytic cysteine is positioned at the N-terminal end of α helix, and the catalytic histidine is part of a β strand. However, organization of the tertiary structure of the Sindbis nsP2/3 protease domain and VEEV nsP2 protease domain is different from that observed in papain and any other known protein structure.

The protease domain of Sindbis nsP2 consists of seven α helices with two regions of short β hairpins. Based on the DALI (Holm and Sander, 1995), the structure of nsP2/3 was very similar to the structure of the VEEV nsP2 protease domain by superposition using the Pymol program. With the exception of this protease domain, the structure of nsP2/3 shows very low similarity to several cysteine proteases, including papain, several cathepsins, and the ubiquitin carboxyl terminal hydrolase. The root mean square deviation (RMSD) between the Sindbis nsP2/3 protease domain and the VEEV nsP2 is 0.987. The DALI program identified the VEEV nsP2 protease (PDB code: 2hwk) as having a high similarity to the Sindbis nsP2/3 protease domain, with a z score of 22.8 and 47% identity. Also the DALI program identified papain like protease inhibitor (PDB code: 3mj5) and cathepsin B (PDB code: 1gmy) as having little structural similarity to the nsP2protrease domain. Both the Sindbis nsP2/3 protease domain and the VEEV nsP2 protease domain have limited similarity in structural alignment.

The Sindbis nsP2/3 protease domain is identified as having low similarity to the cysteine proteinase superfamily (SCOP). The structure of these cysteine proteinases

consists of a common catalytic core made of one α helix and three strands of β sheet (Murzin et al., 1995). Among the cysteine protease structures, the Sindbis nsP2/3 protease domain, the VEEV nsP2 protease domain and papain like protease, the catalytic cysteine is situated at the N terminus of a helix, with the catalytic histidine located on a β strand. In papain like protease domain, the cysteine and histidine residues of the catalytic dyad are located in two separate domains, and the active site occupies the interface between these domains (Drenth et al., 1968). In contrast, in Sindbis nsP2/3 and VEEV nsP2 protease, the β strand containing the catalytic histidine is part of a short β hairpin within the N-terminal protease domain. The structure of Sindbis nsP2/3 supports the hypothesis, derived from mutagenic analysis (Strauss et al., 1992), which suggests a catalytic dyad within the protease active site is composed of Cys481 and His558. A conserved asparagine (Asn561) is located near these residues, and could potentially function as the third element of a catalytic triad. However, the structure shows that this asparagine residue is not oriented to interact with the histidine residue. Moreover, mutational studies of the related Sindbis virus (SINV) show that this residue is not essential for activity (Strauss et al., 1992). In addition, Trp559Ala is completely inactive, suggesting that Trp559 is essential for a functional proteinase. This trp559 is located in a β strand and close to Cys481 and His 558.

3.4 Sindbis nsP2 methyltransferase like domain (MT like domain)

Sindbis nsP2 MT like domain extends from Arg615 to Ala807. This domain contains a central twisted six stranded β sheets surrounded by 4 helices, two on one side and two on the other (Fig. 9).

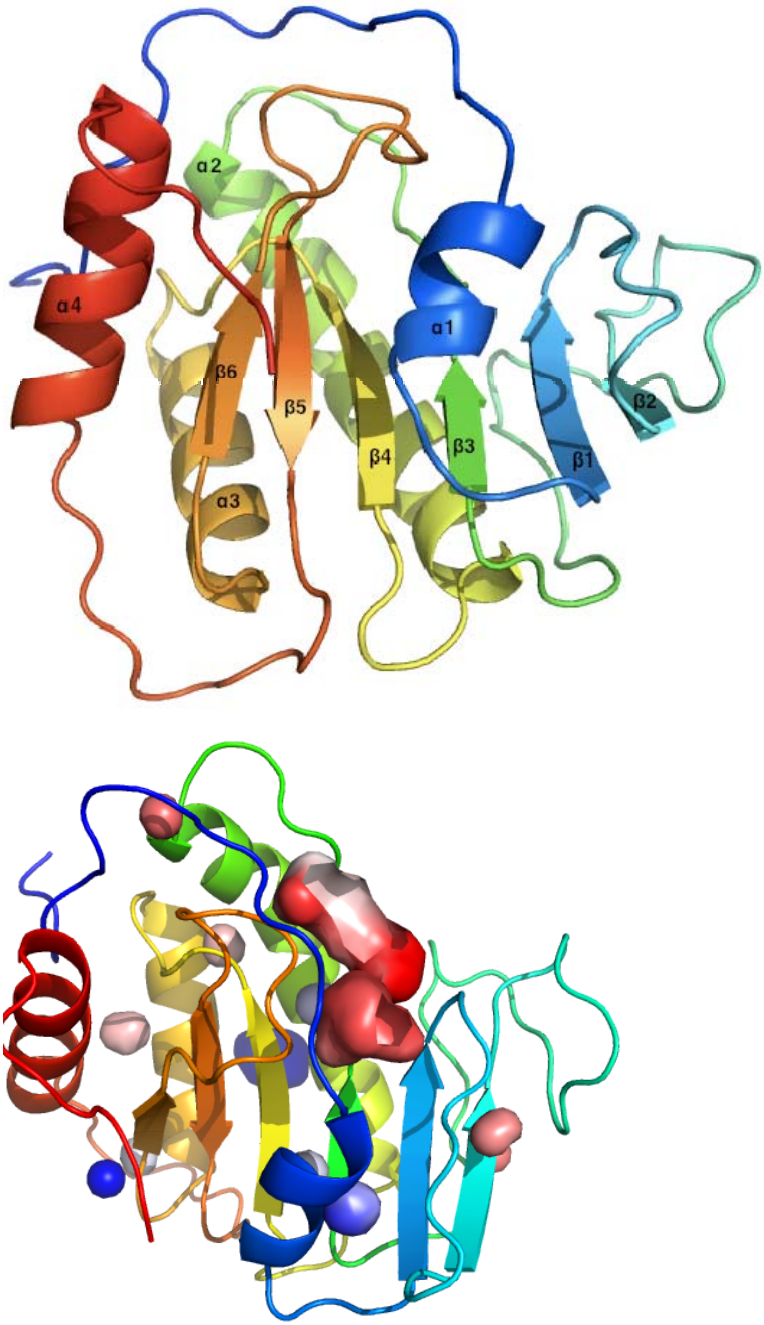


Figure 9. Structure of Sindbis nsP2 methyltransferase like domain shown in rainbow. Structural similarity between S-adenosyl-L-methionine-dependent RNA methyltransferase and methyltransferase-like domain

The function of the MT like domain in nsP2/3 protein is not well known**,** but based on the Dali program (Holm and Sander, 1995), all data indicate that the tertiary structure of the second domain of Sindbis nsP2/3 has high similarity to the structure of the VEEV nsP2 C-terminal domain with a z-score of 26.7%. Both the Sindbis nsP2/3 second domain and the VEEV nsP2 C-terminal domain are similar to the methyltransferase family of enzymes (Russo et al., 2006). Although the Sindbis nsP2/3 second domain has a very similar structure to the VEEVnsP2 domain, the main structural difference is in the connection between the β2 sheet and β3 sheet. The RMSD between the Sindbis nsP2/3 MT like domain and the VEEV nsP2 MT like domain is 0.667. In the case of the VEEV nsP2 domain, it contains an extra short β sheet and α helix between the β2 sheet and β3 sheet. The MT like domain in Sindbis nsP2/3 has a significant tertiary structure which is similar to proteins belonging to the S-adenosyl- L-methionine (SAM)-dependent methyltransferases superfamily. This family includes methyltransferases from a wide variety of organisms, including RNA polymerase NS5 and the *E. coli* heat shock protein FtsJ RNA methyltransferase. However, S-adenosyl-L-methionine binding site between Sindbis nsP2/3 MT like domain and FtsJ methyltransferase domain showed low similarity among individual residue and the overall structural alignment was very poor in binding site. Based on Dali (Holm and Sander, 1995), there is a 38% sequence identity between the Sindbis nsP2/3 MT like domain and the VEEV nsP2 C-terminal domain. Also, this domain has 13% sequence identity with the ribosomal RNA large subunit methyltransferase J, ribosomal RNA methyltransferase and the *E. coli* heat shock protein FtsJ RNA methyltransferase. In a previous study (Sawicki et al., 2006), sequence similarity between alphavirus nsP2 and 2'-O-methyltransferase (2'-O-MT) suggests that this domain is

enzymatically inactive. The common core SAM-dependent methyltransferase superfamily fold was described as having seven stranded β sheets and three helices on each side as sandwiched form. The β sheets of seven strands are arranged in the order 3-2-1-4-5-7-6 with antiparallel to the other strands (Jennifer and Fiona 2002). One of the α helix layers consists of a single helix that is 5 residues long. The Sindbis nsP2/3 second domain has a tertiary structure similar to SAM dependent methyltransferase structures (e.g., VEEV nsP2 C-terminal domain, FtsJ, and RNA polymerase NS5). All of the β strands between Sindbis nsP/2 and FtsJ(PDB:1ejo) align well. However, the backbone alignment, when compare to SAM is poor. The SAM substrate binding sites show no significant similarity to each other.

3.5 Sindbis nsP3 Macro domain

The structure of the Sindbis nsP3 macro domain consists of a central twisted six-stranded β sheet surrounded by three helices on one side and three on the other (Fig. 10).

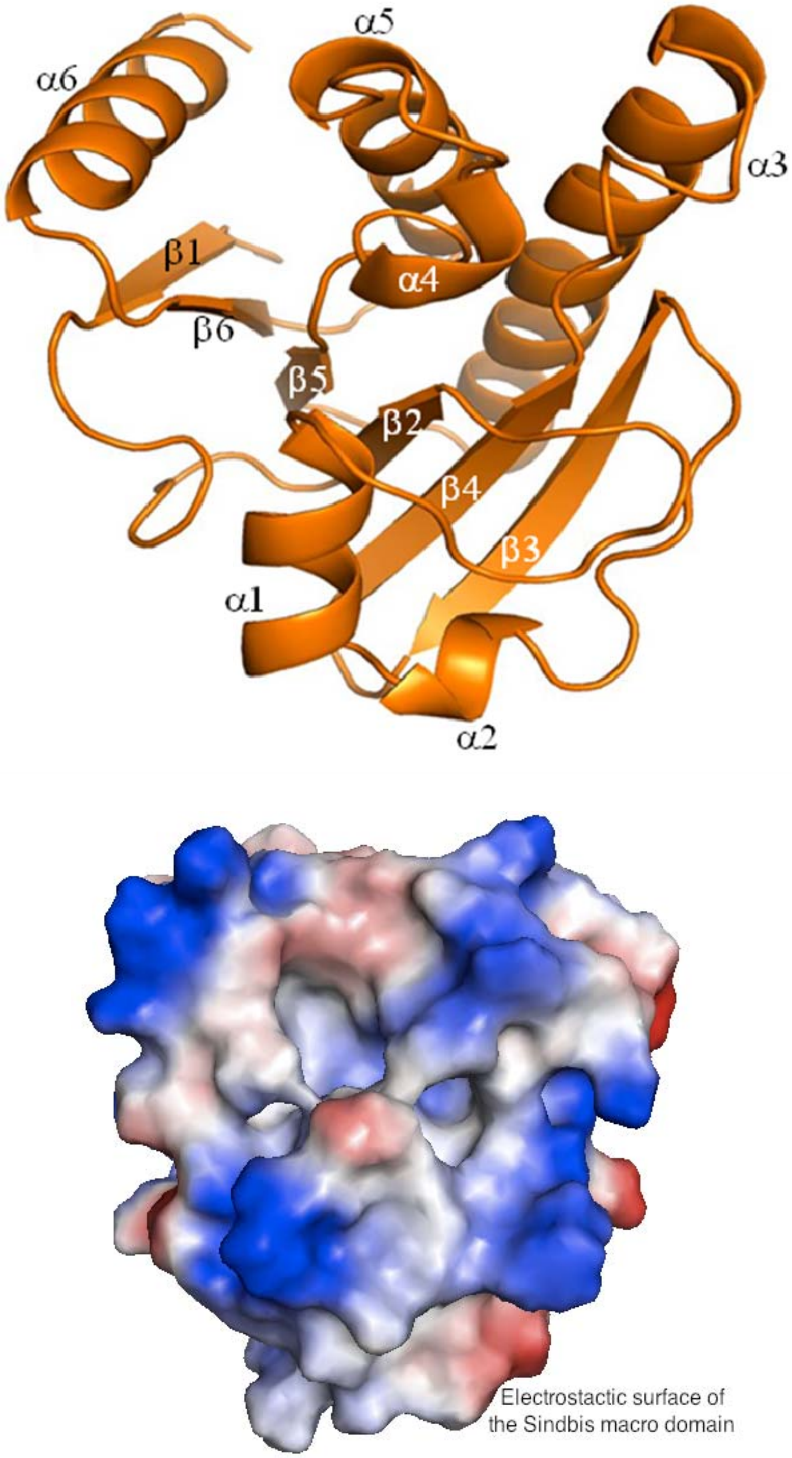


Figure 10. Structure of Sindbis virus nsP3 macro domain shown in orange. It is known that the macro domain binds to ADP-ribose. Electrostatic surface of Sindbis macro domain.

Based on previous structural studies of macro domain in alphavirus, the core β sheet and the positions of the α helices are well conserved within the structures of macro domains. Sequence identity between 160 amino acid residues of the Sindbis nsP3 macro domain and the CHIKV macro domain was very high with 57% identity and z-score of 31. Apparently, the most closely related structures to the Sindbis nsP3 macro domain are the CHIKV and VEEV nsP3 macro domain as well as the putative phosphatase of *Escherichia coli* macro domain. The RMSDs among the alphavirus macro domains are between 0.8 and 1.1 Å for 160 amino acids. When searching the Dali server (Holm and Sander, 1995), the Sindbis nsP3 macro domain had significant structural and functional similarity to a putative poly protein of *E.coli*, core histone macro H2A, and the replication polyprotein of SARS-CoV macro domain as well as other alphavirus macro domains such as CHIVK and VEEV macro domain (Ratia et al., 2006; Saikatendu et al., 2005).

Interestingly, the active site of the Sindbis nsP3 macro domain contains 2-(N-morpholino) ethane sulfonic acid (MES) binding pocket, which may function as a substrate-binding site. This MES binding cleft matches with most conserved active site residue. The highly conserved residues surrounding the buffer molecule suggest that the MES is mimicking a natural ligand. In addition, glycin rich conserved residues in the active site are solvent exposed.

The binding pocket of the macro domain was similarly defined as an ADP-ribose 1"phosphatase active site (Karass et al., 2005). The ADP ribose-binding site is located in a deep hydrophobic pocket, which is on the upper side of the β2, β4 and β5 near the by α1- β2 and α3- β5 connecting loops. Previous structural studies of other macro domains proved that the adenine moiety is selectively H-bonded by its N6 nitrogen to the Asp817 side chain

(Malet et al., 2009; Karras et al., 2005). Although the Asp817 residue is conserved in most macro domain sequences, only Sindbis virus macro domain has an Asn817 residue that could provide a carbonyl group to form a hydrogen bond with the N6 of an adenine.

Electrostatic surface analysis of the Sindbis nsP3 macro domains shows a highly positively charged patch located both in the crevice, (previously known as an ADP-ribose 1-phosphate phosphatase active site) and at its periphery. The opposite site of the nsP3 macro domain, located far from the active site, is negatively charged. The positive and negative charge distribution is more commonly found in alphavirus macro domains than in other macro domain structures (Malet et al, 2009).

Residues near the active site in the macro domain define a positively charged pocket, which is likely to bind a bulky negatively charged group. Conserved positive patches on the electrostatic surface of the protein suggest possible binding site for longer negatively charged chains such as PAR or RNA. Studies have shown that the alphavirus macro domains show stronger PAR binding than SARS-CoV macro domain (Malet et al., 2009; Kumar et al., 2005). Neuvonen and Ahola (2009) proved that several viral macro domains could bind RNA. Those studies suggested that alphavirus macro domain could function in viral RNA replication and transcription.

In addition, there is a channel in the macro domain near the linker connecting the macro domain and zinc binding domain. Interestingly, this channel is positively charged hole on one side and negatively charged hole on the other. Moreover, the structure of the macro domain alone does not contain this channel at all. This is totally new observation that appears to provide a nice binding pocket, possibly functioning in the viral polyprotein processing (Fig. 7).

3.6 Sindbis nsP3 Zinc binding domain

Zinc binding domain begins at residue Glu967 (Glu1507) and continues to the end of the sequence at Tyr1132 (Tyr1672). The 48 amino acids in length begin right after the macro domain, connecting it to the zinc binding domain (Fig. 11).

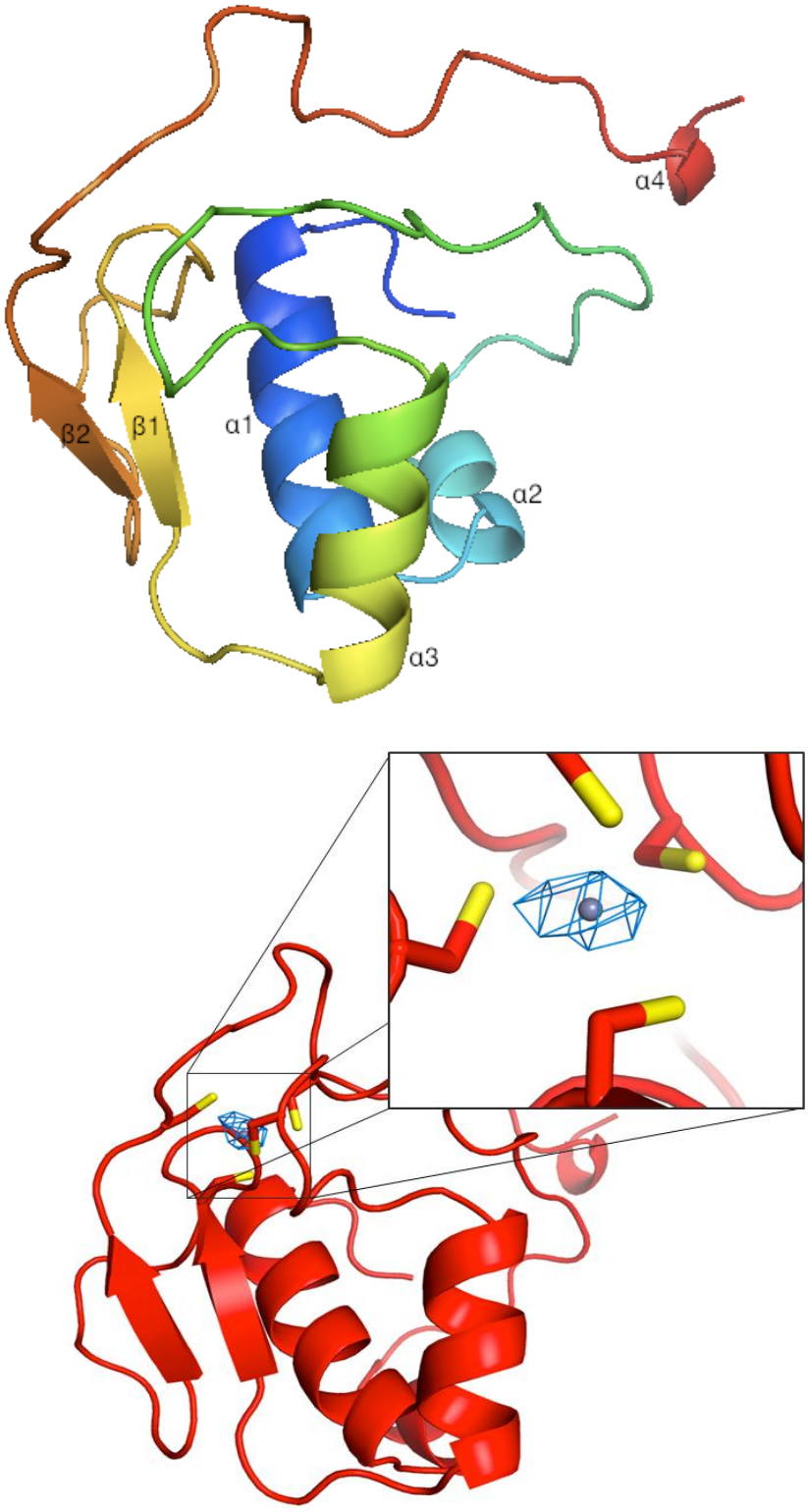


Figure 11. Structure of Sindbis virus nsP3 zinc binding domain shown in rainbow(upper) and red(bottom). The nsP3 contains a novel fold with a previously

uncharacterized zinc coordination site. Electron density of the zinc atom corresponding to 8σ is shown.

The C-terminal domain of the Sindbis nsP2/3 protein turned out to be a zinc binding domain. The phasing was determined through a combination of molecular replacement and single wavelength anomalous dispersion from the zinc. The zinc binding domain consists of four α helices and two stranded β sheets. The zinc ion is located between two loops. One connecting the α2 helix and α3 helix and the other connecting the β2 sheet and α4 helix. Four cysteine residues are involved in the zinc coordination. The distances observed between the cysteine side-chain sulphur groups and the zinc atom are as follows: Cys1610 (2.63 Å), Cys1612 (2.69 Å), Cys1635 (2.69 Å) and Cys1653 (2.54 Å). All cysteine residues are absolutely conserved among all known alphavirus. Based on structural classification of zinc fingers (Krishna et al., 2003), Sindbis nsP3 zinc binding domain is possibly categorized as a short zinc binding loops in fold group 7. Commonly, the zinc-binding domain in this group has at least three zinc ligands that are closely located within the sequence and the zinc ion could stabilize loops. Using the DALI server (Holm and Sander, 1995), no structures were found related to the Sindbis nsP2/3 protein, indicating that this protein contains a novel fold with a previously uncharacterized zinc coordination site. However, the biological function of this zinc binding domain is still unknown and would provide an interesting topic of study.

The linker connecting the macro domain and zinc binding domain starts at residue Glu967 (Glu1507) and end at Phe1014 (Phe1554). This linker directly contacts the long α2 helix in the MT like domain. When the Sindbis nsP2/3 structure was superimposed

with the VEEV nsP2, the structure of the MT like domain was shifted down due to the contact between the connecting loop and the α helices in MT like domain. However the structure of the VEEV nsP2 protease domain did not shift down and aligned well with the structure of Sindbis nsP2 protease domain (Fig. 6A and Fig. 6B).

Electrostatic surface analysis of the Sindbis nsP2/3 shows that the macro domain contains/zinc binding domain linker contains almost half highly negative charged from residue Glu967 (Glu1507) to residue Glu982 (Glu1522) and half highly positive charged from residue Leu983 (Leu1523) to residue Phe1014 (Phe1554). Interestingly, temperature sensitive mutations in the MT like domain, Pro 726 (Pro1266) and Arg751 (Arg1291), are located in very close proximity to this loop region (Fig. 12).

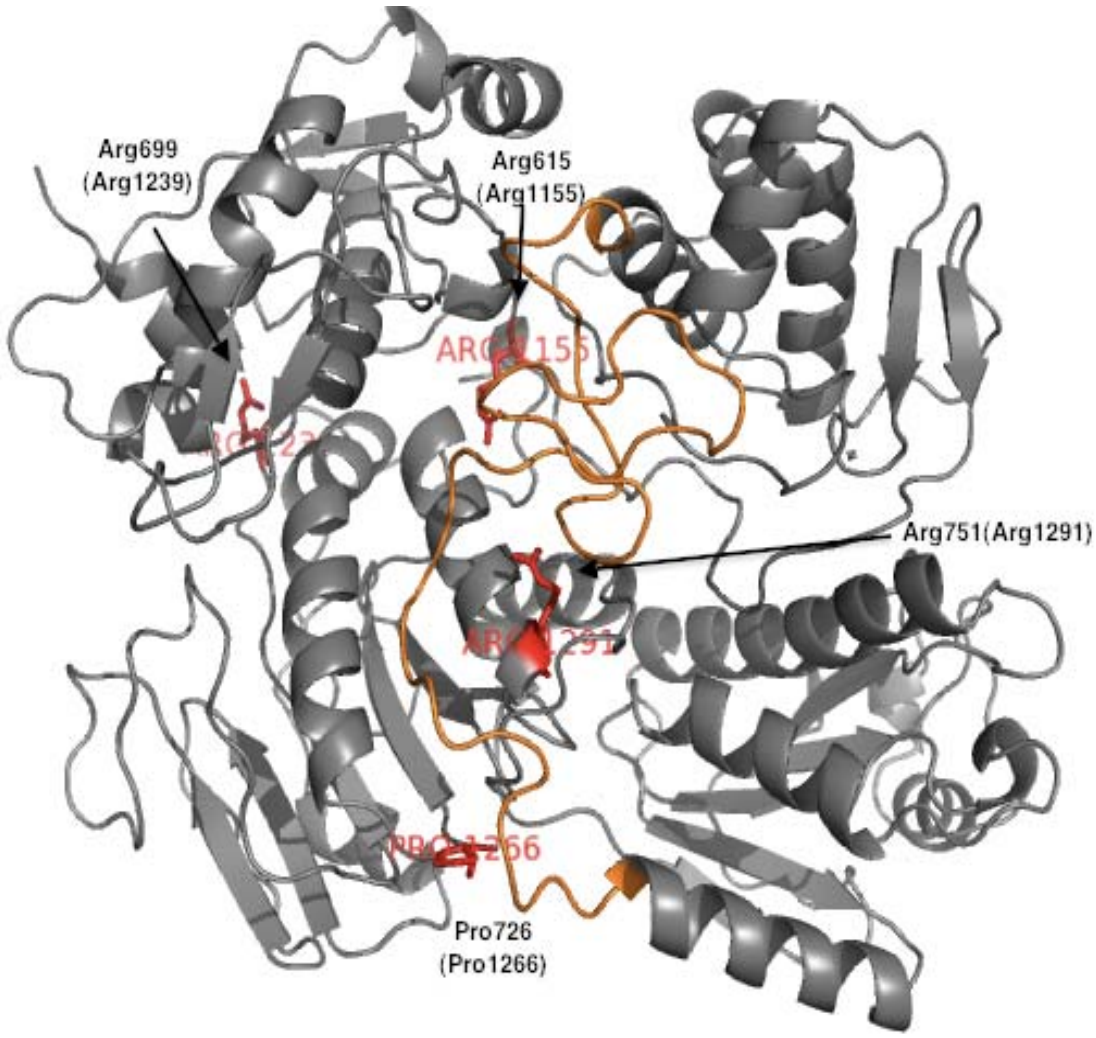


Figure 12. Location of temperature Sensitive Mutants is shown onto the Sindbis nsP2/3 structure. Arg615 hides in the middle of nsP2/3 structure. Arg 615 contributed to efficient plus strand RNA replication. Arg751 and Pro726 are toward to the loop connected between the macro domain and the zinc binding domain. Pro726 is known to reduce the viral cytopathic effect and replication in cell culture.

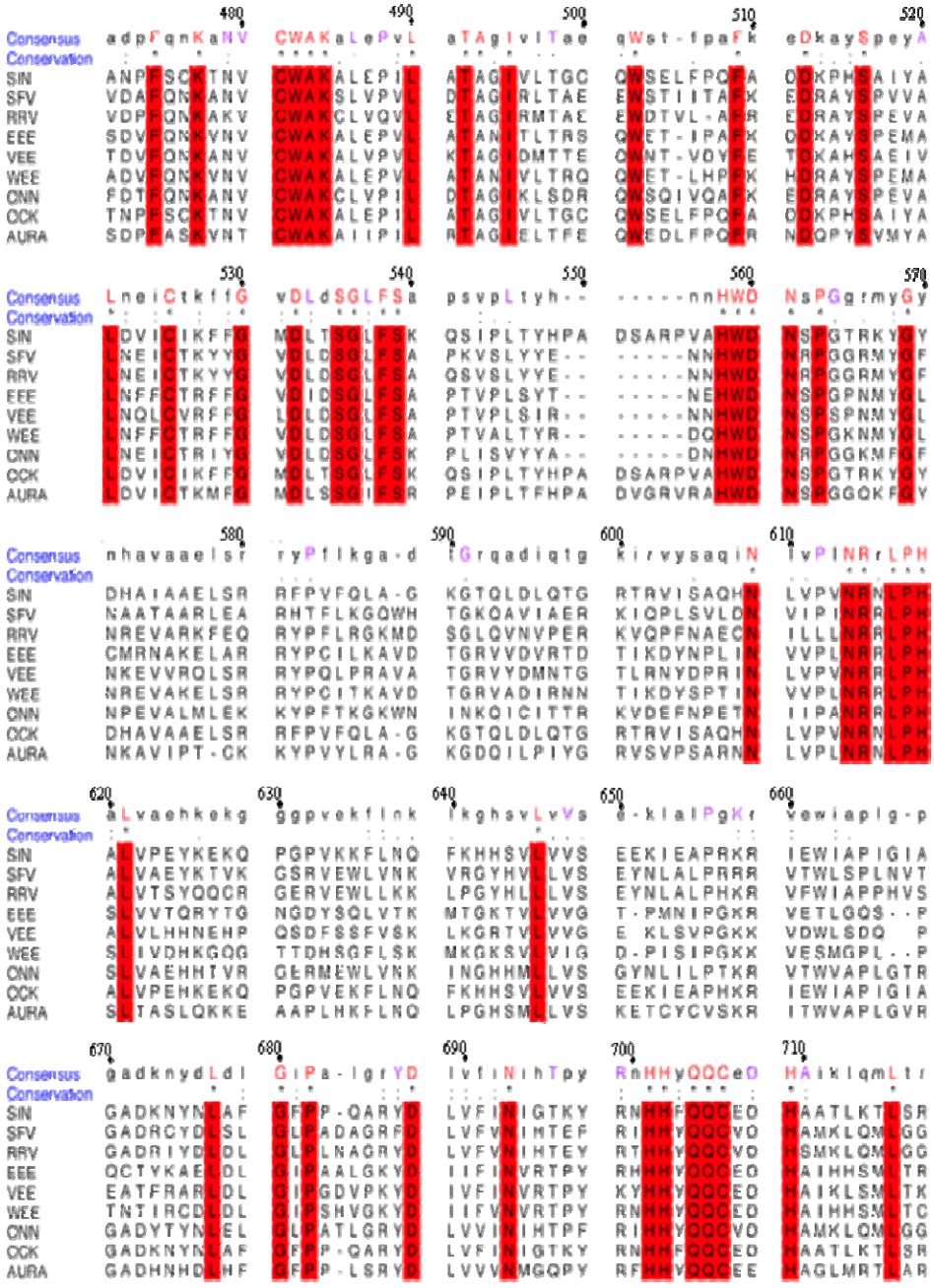
3.7 Functional mutations in Alphavirus nsP2/3

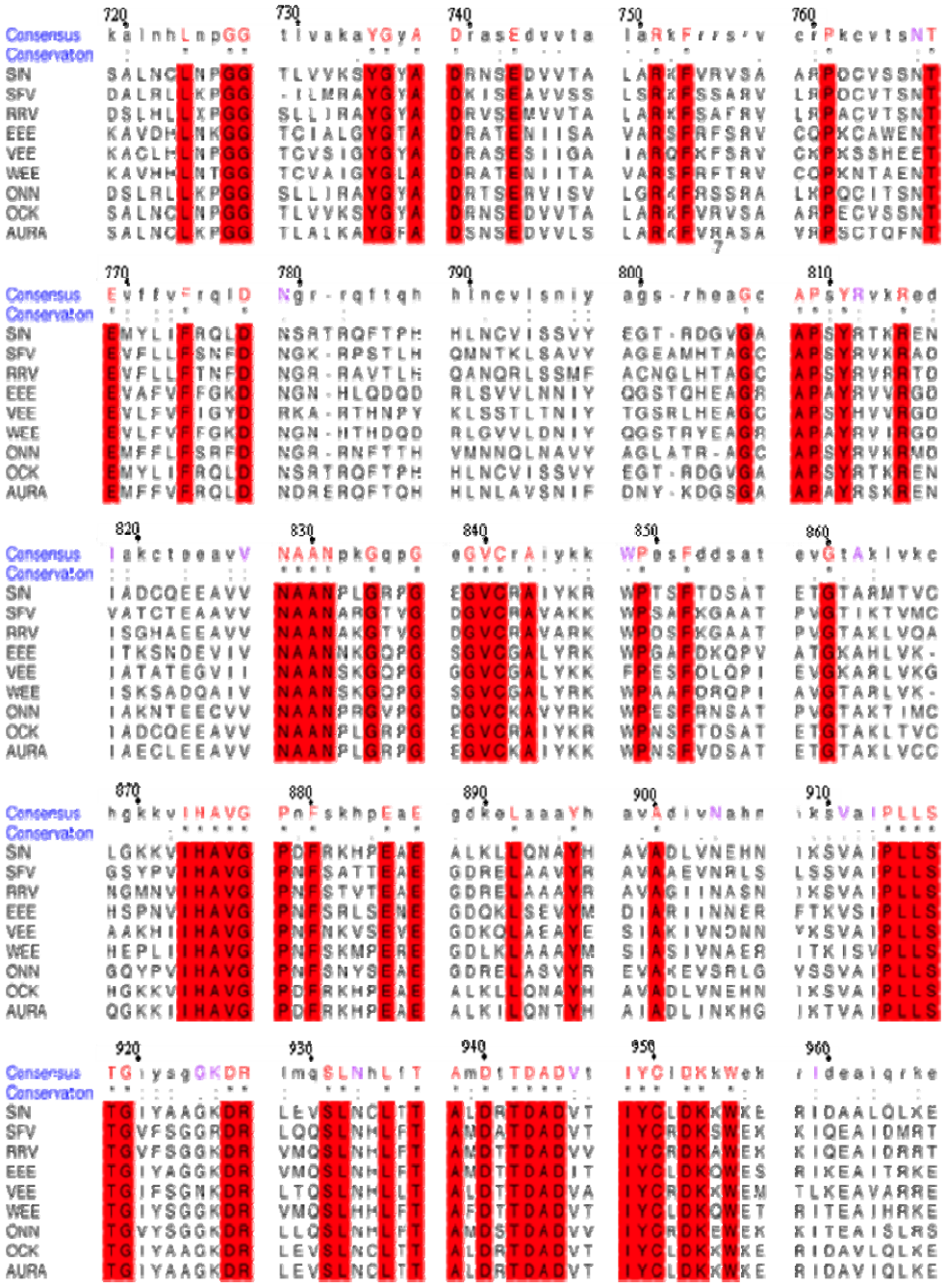
A number of temperature sensitive (ts) mutations in the protease and the MT like domains in nsP2 have been identified which cause defects in subgenomic RNA synthesis, defects in protease activity and cessation of minus strand synthesis in Sindbis virus, Semlik forest and Venezuelan Equine Encephalitis alphavirus (Russo et al., 2006; Hardy et al., 1990; Suopanki et al., 1998; Sawicki et al., 2006). Importantly, temperature sensitive mutations have been shown to be very well conserved among alphavirus strains. The temperature sensitive mutation, ts24 (G723S) of glycine to serine in VEEV MT like domain results in defects both protease activity and 26S RNA synthesis at nonpermissive temperatures. In the structure of Sindbis nsP2, the glycine residue at ts24 (G736S) is completely buried and surrounded by the side chains of conserved residues His619, Tyr798, Tyr737and Ala738. Residue R615 in Sindbis nsP2 plays an important role for efficient plus strand RNA replication. Mutation of R615 causes specific defects in the protease activity of nsP2 and results in inhibition of host translation, during minus strand RNA synthesis (Mayuri et al., 2008). In addition, the adaptive mutation of MT like domain at residue P726 of Sindbis nsP2 (P713 VEEV nsP2) has been reported to reduce the viral cytopathic effect and replication in cell culture in a cell line dependent manner (Frolova et al., 2002). Pro726 is found in a tight turn at the end of a helix and participates in a parallel special β bulge (Russo et al., 2006).

In Sindbis macro domain nsP3, highly temperature sensitive mutations CR3.36 and CR3.39 are defective in their ability to synthesize minus-strand RNAs at the nonpermissive temperature (LaStarza et al., 1994; Lemm and Rice, 1994). Mutations of asparagine residue 817 and 831 to alanine in the ADP-ribose binding region of the nsP3

macro domain impaired Sindbis virus replication and viral RNA synthesis for neuronal study. This mutation study of the nsP3 macro domain at amino acid 10 and 24 showed impaired virus replication, impaired viral RNA amplification, and induced instability of nsP3 in neurons (Park and Griffin, 2009).

Starting number of Sindbis Sequence is Ala 471 (Ala 1011)




720 730 740 750 760
Consensus
Conservation
SIN
SFV
RRV
EEE
VEE
WEE
ONN
OCK
AURA
770 780 790 800 810
820 830 840 850 860
870 880 890 900 910
920 930 940 950 960

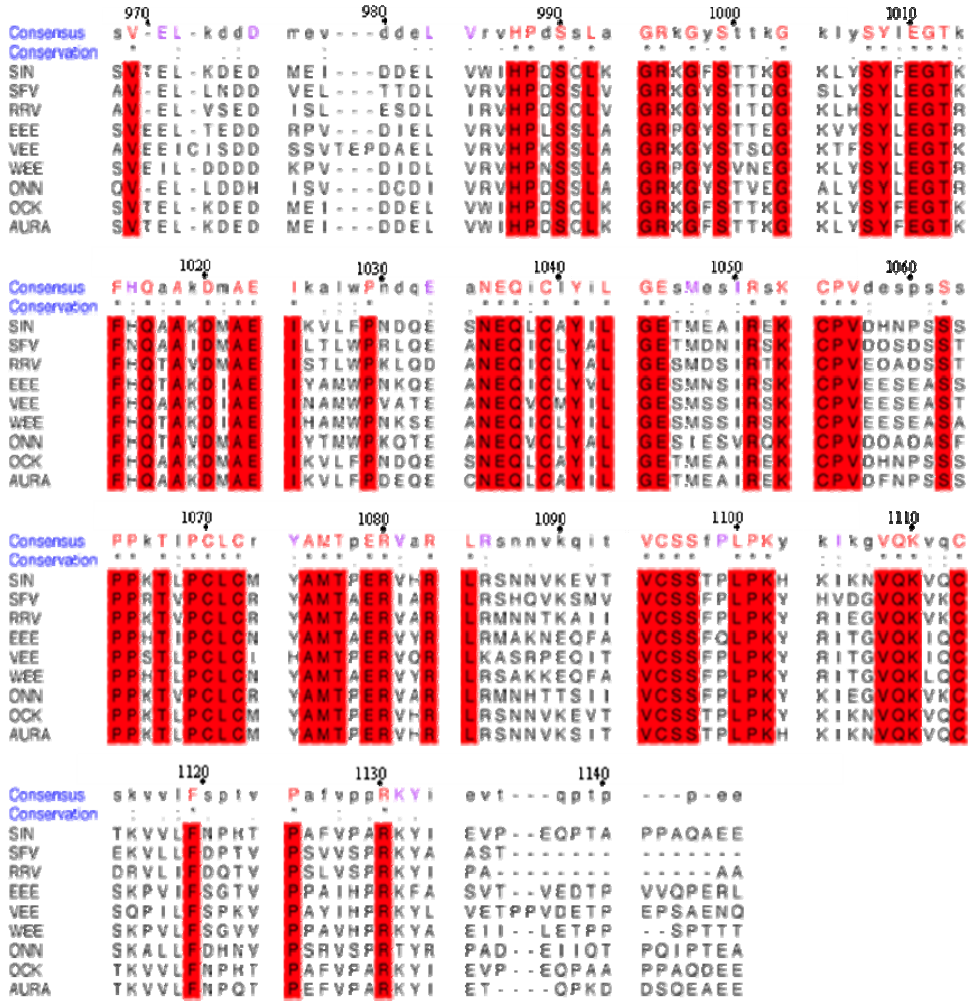


Figure 13. Sequence alignment of Alphaviruses. Strictly conserved residues are boxed in red.

Chapter 4

Discussion

Proteolytic sequential processing of the alphavirus polyprotein is still not well understood. The nsP3/4 site can be cleaved *in trans* by the nsP2 protease C-terminal domain, and then the nsP1/2 junction is cleaved *in cis* by the same protease. However, it is not known how the nsP2/3 junction is recognized and cleaved by the nsP2 protease. There is evidence which shows that the nsP2/3 cleavage site can not be recognized by only the nsP2 protease C-terminal domain (Vasiljeva et al., 2003). This study proved that the amino terminus of nsP2 is necessary to cleave the nsP2/3 site in trans. Based on our structural data, the cleavage sites for nsP2 and nsP3 are buried in the two domains, making them inaccessible for proteolysis. It is possible that structure might requires a conformational change to enable cleavage site access for proteolytic processing. A structure of the Sindbis nsP2/3 precleavage form plays an important role for further studies of alphavirus polyprotein proteolytic processing and viral replication.

Recently, the functional importance of MT like domain within the nsP2 C-terminal region has been discussed. Several important temperature sensitive mutations are located within the MT like domain. The S2 mutation (Pro726) in Sindbis virus is known to reduce the viral cytopathic effect and results in persistent infection of Sindbis virus. Residue Pro726 is found at the end of a α2 helix within the structure and toward the loop region. Myuri et al., (2008) proved that mutation of residue Arg615 in Sindbis nsP2 MT like domain had the viral noncytopathic effect and could not inhibit host cell translation. Based on our observations, this Arg615 residue is located in the linker region between the protease domain and the MT like domain and hides in the center of overall Sindbis nsP2/3

structure to difficult to access this residue. It has been proposed that this linker region plays an important role in forming a functional substrate binding pocket for the protease between the protease domain and MT like domain (Myuri et al., 2008).

Based on previous studies (Malet et al., 2009), the nsP3 macro domain of alphaviruses exhibited a higher affinity for binding ADP-ribose, ADP-ribose 1"phosphate phosphatase activity and PAR binding activity. Based on the structure of the macro domain, the ADP- ribose binding pocket in the macro domain is a highly positively charged region. This positively charged binding pocket is well conserved among alphaviruses and shows possible binding of negatively charged molecules and even longer molecules such as PAR and RNA.

Interestingly, Sindbis nsP3 macro domain has a channel next to the negatively charged linker connecting the macro domain and zinc binding domain. This channel is positively charged on one side and negatively charged on the other side. This novel channel observed in the Sindbis nsP2/3 structure did not exist in the structure of the macro domain only. This channel has a nice binding pocket and may function in the viral polyprotein processing.

Structural comparisons using DALI (Holm and Sander, 1995) and Protein Data Bank showed that zinc binding domain of Sindbis nsP3 had no similarity with any other zinc binding proteins known to this point. It indicates that Sindbis nsP3 contains a novel fold with a previously uncharacterized zinc coordination site. We found a new zinc binding domain in Sindbis nsP3 protein, although its biological function of zinc binding domain is still unknown.

Another interesting observation within the structure is the linker between the macro domain and the zinc binding domain. The half negatively and half positively charged long linker directly contacts α helices within the MT like domain. It causes the structure of MT like domain to shift down without changes in any other domain. This long linker runs right through the middle of the overall structure. Important temperature sensitive mutants such as Pro726 (Pro1266) and Arg751 (Arg1291) is located nearby this linker region. The mutation in residue Pro726 (Pro1266) disrupts nsP2-mediated viral RNA synthesis (Sawicki et al., 2006). In addition, it is known that this mutation could reduce the viral cytopathic effect and replication in Sindbis virus and Semliki forest virus (Frolova et al., 2002). Based on the observation of the linker region position and a couple of mutants, this linker region is likely to play an important structural role in viral RNA synthesis and viral replication.

References

Ahola T, Kaariainen L. 1995. Reaction in alphavirus mRNA capping: formation of a covalent complex of nonstructural protein nsP1 with 7-methyl- GMP. Proc. Natl. Acad. Sci. 92:507–511

Drenth J, Jansonius JN, Koekoek R, Swen HM, Wolthers BG. 1968. Structure of papain. Nature 218:929–932.

Frolova EI, Fayzulin RZ, Cook SH, Griffin DE, Rice CM, Frolov I. 2002. Roles of nonstructural protein nsP2 and Alpha/Beta interferons in determining the outcome of Sindbis virus infection. J. Virol. 76:11254–11264

Gomez DC, Ehsani MN, Mikkola ML, Garcia JA, Kaariainen L. 1999. RNA helicase activity of Semliki Forest virus replicase protein NSP2. FEBS Lett. 448:19–22

Gorchakov R, Frolova E, Sawicki S, Atasheva S, Sawicki D, Frolov I. 2008. A new role for ns polyprotein cleavage in Sindbis virus replication. J. Virol. 82(13):6218–6231

Hardy WR, Hahn YS, de Groot RJ, Strauss EG, Strauss JH. 1990. Synthesis and processing of the nonstructural polyproteins of several temperature-sensitive mutants of Sindbis virus. Virology 177:199–208.

Holm L, Sander C. 1995. Dali: a network tool for protein structure comparison. Trends Biochem. Sci. 20:478–480

Jennifer LM, Fiona MM. 2002. SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. Curr. Opin. Struc.Biol. 12:783-793

Karras GI, Kustatscher G, Buhecha HR, Allen MD, Pugieux C, Sait F, Bycroft M, Ladurner AG. 2005. The macro domain is an ADP- ribose binding module. EMBO J. 24:1911–1920

Koonin EV, Gorbalenya AE, Purdy MA, Rozanov MN, Reyes GR, Bradley DW. 1992. Computer-assisted assignment of functional domains in the nonstructural polyprotein of hepatitis E virus: delineation of an additional group of positive-strand RNA plant and animal viruses. Proc. Natl. Acad. Sci. USA. 89:8259–8263.

Krishna SS, Majumdar I, Grishin NV. 2003. Structural classification of zinc fingers: survey and summary. Nucleic Acids Res. 31: 532–550

Laine M, Luukkainen R, Toivanen A. 2004. Sindbis viruses and other alphaviruses as cause of human arthritic disease. J. Intern. Med. 256:457–471

LaStarza MW, Lemm JA, Rice CM.1994, Genetic analysis of the nsP3 region of Sindbis virus: evidence for roles in minus-strand and subgenomic RNA synthesis. J. Virol. 68:5781–5791.

Lemm JA, Rice CM. 1993. Roles of nonstructural polyproteins and cleavage products in regulating Sindbis virus RNA replication and transcription. J. Virol. 64(4): 1916–1926

Lemm JA, Rumenapf T, Strauss EG, Strauss JH, Rice CM. 1994. Polypeptide requirements for assembly of functional Sindbis virus replication complexes: a model for the temporal regulation of minus and plus-strand RNA synthesis. EMBO J. 13:2925–2934.

Malet H, Coutard B, Jamal S, Dutartre H, Papageorgiou N, Neuvonen M, Ahola T, Forrester N, Gould EA, Lafitte D, Ferron F, Lescar J, de Lamballerie X, Canard B. 2009. The Crystal Structures of Chikungunya and Venezuelan Equine Encephalitis Virus nsP3 Macro Domains Define a Conserved Adenosine Binding Pocket. J. virol. 83(13):6534–6545

Mayuri, Geders TW, Smith JL, Kuhn RJ. 2008. Role for Conserved Residues of Sindbis Virus Nonstructural Protein 2 Methyltransferase-Like Domain in Regulation of Minus-Strand Synthesis and Development of Cytopathic Infection. J. Virol. 82:7284–7297

Murzin AG, Brenner SE, Hubbard T, Chothia C. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. 247:536–540.

Neuvonen M, Ahola T. 2009. Differential activities of cellular and viral macro domain proteins in binding of ADP-ribose metabolites. J. Mol. Biol. 385:212–225

Park E. and Griffin DE. 2009. The nsP3 Macro Domain is Important for Sindbis Virus Replication in Neurons and Neurovirulence in Mice. Virology 388(2): 305–314

Pehrson JR, Fried VA. 1992. MacroH2A, a core histone containing a large nonhistone region. Science 257:1398–1400

Peränen J, Rikkonen M, Liljeström P, Kääriänen L. 1990. Nuclear localization of Semliki Forest virus-specific nonstructural protein nsP2. J. Virol. 64:1888-1896

Peränen J, Laakkonen P, Hyvönen M, Kääriäinen L. 1995. The alphavirus replicase protein nsP1 is membrane-associated and has affinity to endocytic organelles. Virology 208(2):610–620

Ratia K, Saikatendu KS, Santarsiero BD, Barretto N, Baker SC, Stevens RC, Mesecar AD. 2006. Severe acute respiratory syndrome coronavirus papain-like protease: Structure of a viral deubiquitinating enzyme. Proc. Natl. Acad. Sci. USA, 103:5717–5722

Russo AT, White MA, Watowich SJ, 2006. The Crystal Structure of the Venezuelan Equine Encephalitis Alphavirus nsP2 Protease. Structure. 14:1449–1458

Saikatendu KS, Joseph SJ, Subramanian V, Clayton T, Griffith M, Moy K, Velasquez J, Neuman BW, Buchmeier MJ, Stevens RC, Kuhn P. 2005. Structural Basis of Severe Acute Respiratory Syndrome Coronavirus ADP-Ribose-100-Phosphate Dephosphorylation by a Conserved Domain of nsP3. Structure 13:1665–1675.

Sawicki DL, Perri S, Polo JM, Sawicki SG, 2006. Role for nsP2 proteins in the cessation of alphavirus minus-strand synthesis by host cells. J. Virol. 80:360-371.

Shirako Y, Strauss JH. 1994. Regulation of Sindbis virus RNA replication: uncleaved P123 and nsP4 function in minus-strand RNA synthesis whereas cleaved products from P123 are required for efficient plus strand RNA synthesis. J. Virol. 185:1874-1885.

Strauss EG, De Groot RJ, Levinson R, Strauss, JH. 1992. Identification of the active site residues in the nsP2 proteinase of Sindbis virus. Virology 191:932-940.

Strauss JH, Strauss EG. 1994. The alphaviruses: gene expression, replication, and evolution. Microbiol. Rev. 58: 491-562.

Suopanki JD, Sawicki L, Sawicki SG, Kaariainen L. 1998. Regulation of alphavirus 26S mRNA transcription by replicase component nsP2. J. Gen. Virol. 79:309–319.

Taylor RM, Hurlbut HS, Work TH, Kingsbury JR, Frothingham TE. 1955. Sindbis virus: A newly recognized arthropod-transmitted virus. Am. J. Trop. Med. Hyg. 4: 844-846.

Tomar S, Hardy RW, Smith JL, Kuhn RJ. 2006. Catalytic core of alphavirus nonstructural protein nsP4 possesses terminal adenylyltransferase activity. J. Virol. 80:9962–9969.

Vasiljeva L, Merits A, Auvinen P, Kaariainen L. 2000. Identification of a novel function of the alphavirus capping apparatus. RNA 5’-triphosphatase activity of Nsp2. J. Biol. Chem. 275:17281–17287.

Vasiljeva L, Merits A, Golubtsov A, Sizemskaja V, Kaariainen L, Ahola T. 2003. Regulation of the sequential processing of Semliki Forest virus replicase polyprotein. J. Biol. Chem. 278:41636–41645.

Wang HL, O’Rear J, Stollar V. 1996. Mutagenesis of the Sindbis virus nsP1 protein: effects on methyltransferase activity and viral infectivity. Virology 217:527–531