Structural and biochemical characterization of non-structural protein 3 from flavivirus

Phoo, Wint Wint

2018

Phoo, W. W. (2018). Structural and biochemical characterization of non-structural protein 3 from flavivirus. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/87661>

<https://doi.org/10.32657/10220/46782>

**STRUCTURAL AND BIOCHEMICAL CHARACTERIZATION OF NON-
STRUCTURAL PROTEIN 3 FROM FLAVIVIRUS**

WINT WINT PHOO

School of Biological Sciences

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

1st June 2018

Acknowledgments

May we have the serenity to accept the things we cannot change; courage to change the things we can; and wisdom to know the difference.

Serenity Prayer

First and foremost, I thank my supervisor, Asst. Prof. Luo Dahai for his guidance and support throughout my PhD journey and for teaching how to conduct scientific research from experimental designs to the publications. Without his help and support, my graduate studies would not have been possible. I thank my thesis advisory committee, Assoc. Prof. Julien Lescar who is also my co-supervisor and Dr. Siew Peng Lim, for their guidance, fruitful discussions and encouragement during my PhD. I thank Prof. Subhash Vasudevan for his help with not only the experiments but also with the manuscript writing and critical reading. I thank our collaborators, Dr. Congbao Kang and Dr. Li Yi from A*STAR and Prof. Steinmetzer group from University of Margburg, for the prolific collaborations.

I thank all the previous and current lab members for their help along my PhD and for providing me with warm and peaceful research environment. I thank Dr. Ming Wei Chen for helping me with structural biology especially during the early years of PhD. I thank Ms. Zhenzhen Zhang for excellent technical support for the project. I thank my fellow PhD students, Ms. Yong Hui Yee and Ms. Kuohan Li and Dr. Wirawan, for the moral support and kindness they have given me during the ups and the downs of PhD.

Getting through my PhD would not be possible without the support from my social circle. I thank my housemates and my friends, Andrey and Blaz, Ando and Yorick for their friendship and banters. I thank Adrian for his understanding throughout my PhD years. Most importantly, I thank my parents, to whose support, understanding and upbringing, I owe everything that I am now.

Table of contents	
Acknowledgments	2
List of figures	Error! Bookmark not defined.
List of tables	8
Table of abbreviations	9
Summary	12
1. CHAPTER 1: BACKGROUND	13
Chapter 1.1 History, Taxonomy, Pathogenesis	13
Section 1.1.1: <i>Flaviviridae</i> family viruses: overview on pathogenesis, molecular biology and anti-viral strategies	14
Chapter 1.2: Molecular biology of Flavivirus	28
Section 1.2.1: Flavivirus life cycle	28
Section 1.2.2: Structural proteins from flavivirus	40
Section 1.2.3: Non-structural proteins	43
Chapter 1.3: Immune response and evasion strategies by flavivirus	56
Innate immunity	56
Chapter 1.4: Diagnostic methods, antivirals and vaccines for flavivirus infections	63
Chapter 1.5: X-ray crystallography in drug-discovery	71
Chapter 1.6: NS3 protease: An antiviral target	78
Section 1.6.1: General Mechanism of proteases	78
Section 1.6.2: Structural biology of NS2B-NS3 protease	80
Section 1.6.3: Antivirals targeting NS2B-NS3 protease	89
Chapter 1.6: Aim of the thesis	96
2. CHAPTER 2: MATERIALS AND METHODS	98
Chapter 2.1: Molecular cloning and site directed mutagenesis	98
Chapter 2.2: Expression and purification	99
Chapter 2.3: Crystallization	101
Section 2.3.1: Crystal screening	101
Section 2.3.2: Crystal optimization	101
Chapter 2.4: Data collection, Structure determination and refinement	102
Section 2.4.1: Molecular replacement for DENV4 NS3 full length	104
Section 2.4.2: Molecular replacement for ZIKV protease structures	104
Chapter 2.5: Biophysical and Biochemical assays	105
Section 2.5.1: Protease activity assay	105
Section 2.5.2: Protease Inhibition Assay	107
Section 2.5.3: Thermal shift Assays	108
Section 2.5.4: Molecular weight determination by mass spectrometry	109

3. CHAPTER 3: STRUCTURAL AND BIOCHEMICAL CHARACTERIZATION OF ZIKV PROTEASE 112

Chapter 3.1: Structure determination and enzymatic characterization of ZIKV NS3 protease constructs 113

Section 3.1.1: Structural studies of NS3 protease	113
Section 3.1.2: Expression, purification and enzymatic assays of ZIKV NS3 protease constructs	115
Section 3.1.3: Crystal structure of ZIKV NS2B-NS3 after self-cleavage (eZiPro)	119
Section 3.1.4: Crystal structure of bivalent ZIKV protease (bZiPro)	122
Section 3.1.5: Discussion	126

Chapter 3.2: Exploring the substrate binding sites of the Zika protease 129

Section 3.2.1: Crystal structures of bZiPro in complex with fragment molecules	130
Section 3.2.2: Crystal structures of bZiPro in complex with dipeptides inhibitors	141
Section 3.2.3: Crystal structures of bZiPro in complex with P4-P1' peptidomimetic inhibitors	150
Section 3.2.4: Discussion	161

4. CHAPTER 4: STRUCTURAL AND BIOCHEMICAL CHARACTERIZATION OF NS3 FROM DENGUE VIRUS 165

Chapter 4.1: Structural and biochemical characterization of NS3 protease from dengue virus 166

Section 4.1.1: Construct design, cloning and purification of NS3 protease	166
Section 4.1.2: Crystal structure of eNS2B47NS3 Pro	168
Section 4.1.3: Discussion	174

Chapter 4.2: Structural and biochemical characterization of full length NS3 from dengue virus 176

Section 4.2.1: Construct design, cloning, mutagenesis and purification of full length NS3 constructs	177
Section 4.2.2: Effect of L30S F31S mutation on protease activity	181
Section 4.2.3: Full length structures of NS2B ₄₇ NS3 from DENV4 in apo-form and in complex with BPTI	183
Section 4.2.4: Discussion	200

5. CONCLUSIONS AND FUTURE DIRECTIONS 204

Conclusions	204
Future directions	205

6. APPENDIX 1: MULTIPLE SEQUENCE ALIGNMENT OF NS2B AND NS3 FROM FLAVIVIRUS 206

Multiple sequence alignment of non-structural protein 3 from flavivirus	206
Multiple sequence alignment of non-structural protein 2B from flavivirus	208

7. APPENDIX TABLE 1. AMINO ACID SEQUENCES OF THE CONSTRUCTS USED IN CRYSTALLOGRAPHY 209

8. APPENDIX TABLE 2: LIST OF CRYSTAL STRUCTURES REPORTED IN THE THESIS 211

9. REFERENCES 212

List of figures

Figure 1.1	Baltimore classification of viruses depending on their genetic materials.	14
Figure.1.2	Phylogenetic tree based on the NS5 RNA dependent RNA polymerase (RdRp) of Flaviviridae family. The lengths of the lines are directly proportional to genetic distance.	16
Figure 1.3	Genome organization of different families of flavivirus.	20
Figure 1.4:	Phylogenetic tree of members of flavivirus genus based on polyprotein precursor.	22
Figure.1.5.	Geographical distribution of flaviviruses that are important human pathogens.	
Figure adopted from	www.creative-diagnostics.com/Flavivirus.htm .	23
Figure 1.6.	Transmission cycles of flaviviruses.	25
Figure 1.7.	Flavivirus replication cycle. Figure adapted from (Gerold et al., 2017)	30
Figure 1.8.	Mechanisms of viral fusion with the envelope.	34
Figure 1.9.	Mature and immature viral particles of DENV.	38
Figure 1.10.	Genome organization of flavivirus. The secondary structures at the untranslated regions of 5' and 3'.	40
Figure 1.11.	Flavivirus capsid protein structure.	41
Figure 1.12.	Structure of ectodomains from ZIKV E protein.	43
Figure 1.13.	Crystal structure of NS1 dimers from DENV.	45
Figure 1.14.	Membrane topology of DENV NS2B protein.	46
Figure 1.15.	Structures of NS3 protease from HCV and WNV protease.	49
Figure 1.16.	Crystal structures of NS3 helicase domains from flavivirus	51
Figure 1.17.	Full length crystal structures of HCV, MVEV and DENV	54
Figure 1.18.	Full length NS5 structure of DENV3	56
Figure 1.19:	Major PAMPs recognition receptors and the downstream pathways.	57
Figure 1.20.	IFN antagonism strategies by Flavivirus.	62
Figure 1.21.	Work flow of structure determination by X-ray crystallography.	73
Figure 1.22.	Bragg law of diffraction.	74
Figure 1.23.	Ewald sphere construction.	76
Figure 1.24.	Catalytic mechanism proteases	79
Figure 1.25.	Catalytic mechanisms of serine proteases.	80
Figure 1.26	Crystal structure of NS2B-NS3 protease with peptidomimetic inhibitors.	84
Figure 1.27.	NMR studies of NS2B-NS3 protease.	87
Figure 2.1	Crystal optimization of different constructs..	102
Figure 2.2.	Scheme of protease activity assay.	107
Figure 2.3.	ATPase activity assay	110
Figure 3.1.	Construct design and purification of ZIKV NS2B47NS3 protease constructs.	117
Figure 3.2.	The molecular weights of the proteins are determined by Matrix assisted laser desorption ionization – time of flight (MALDI-TOF).	118
Figure 3.3.	Characterization of gZiPro, eZiPro and bZiPro enzymatic activity.	119
Figure 3.4.	Crystal structure of eZiPro	121
Figure 3.5.	NS2B C-terminus TGKR peptide in the active site of eZiPro molecules.	121
Figure 3.6.	Crystal structure of eZiPro in complex with NS2B/NS3 cleavage peptide, TGKR.	121
Figure 3.7.	Two unit cells of bZiPro. One unit cell consists of four bZiPro molecules.	124
Figure 3.8.	Crystal structure of bivalent NS2B NS3 protease	125
Figure 3.9.	(A) Half maximal inhibitory concentration of EN300.	131
Figure 3.10.	Two pyrazole ester derivative compounds, namely compound 1 and compound 5 interact with bZiPro	133
Figure 3.11.	Mass spectrometry analysis of bZiPro before and after incubation with compound 1.	134

Figure 3.12. Protease activity of C143S and WT demonstrating that C143S mutation does not interfere with the protease activity. The error bars represent standard errors.	135
Figure 3.13. Detailed interactions of bZiPro with compound 1	135
Figure 3.14. 1H-15N HSQC NMR spectra of compound 1 and its derivatives.	139
Figure 3.15. Evaluation of the inhibition efficiency of different dipeptide inhibitors against bZiPro using benzyol-Nle-Lys-Lys-Arg-AMC substrate.	141
Figure 3.16. Crystal structure of bZiPro in complex with a dipeptide inhibitor, Acetyl-Lys-Arg-H.	143
Figure 3.17. 1H-15N HSQC spectra of bZiPro titrated with Ac-KR-H and Ac-K-Agmatine. (The NMR experiments are carried out by Kang Lab.)	144
Figure 3.18. Crystal structure of bZiPro in complex with a dipeptide inhibitor, imidazole-Lys-Arg-H (imidazole-KR-H).	145
Figure 3.19. Comparison of the structures of bZiPro-dipeptide inhibitors with bZiPro-KKGE and with eZiPro-TGKR.	147
Figure 3.20: Chemical structures of P4-P1' peptidomimetics inhibitors and inhibition assays.	151
Figure 3.21. Mass spectra of the compounds before and after incubation with bZiPro determined by ESI-MS.	152
Figure 3.22. Crystal structure of bZiPro in complex with MI751.	155
Figure 3.23. Crystal structure of bZiPro in complex with MI1217.	157
Figure 3.24. Crystal structure of bZiPro in complex with MI1229.	159
Figure 3.25. Superimposed structures of peptidic inhibitors in bZiPro, shown with its backbone in cartoon style with the NS2B colored in magenta and NS3 in yellow.	161
Figure 3.26. Surface hydrophobicity representation of ZIKV NS2B-NS3 protease.	163
Figure 4.1. Schematic drawing of NS2B NS3 gene from DENV and construct design of NS2B47NS3 Protease from DENV4.	166
Figure 4.2. Purification of NS2B-NS3 protease from DENV.	167
Figure 4.3. Characterization of enzymatic activity of the DENV protease constructs.	168
Figure 4.4. Pictures of eNS2B47NS3 Pro crystals.	168
Figure 4.5. Crystal structure of eNS2B47NS3 Pro shows novel conformation of NS2B cofactor.	172
Figure 4.6. Electron density map of NS2B co-factor for four eNS2B ₄₇ NS3 Pro molecules.	172
Figure 4.7 NS2B-NS3 protease dimers in one ASU.	173
Figure 4.8. The NS3 loops 151-158 and 111-121 act as hinges during the NS2B-NS3 protease open and closed conformation.	174
Figure 4.9. Surface charge representation of DENV protease during the partially closed conformation	174
Figure 4.10. Construct designs of DENV4 full length constructs.	178
Figure 4.11. Purification of gNS2B47NS3, eNS2B47NS3 and bNS2B47NS3 showed monomeric proteins.	180
Figure 4.12. Overexpression and purification of DENV4 proteases, gNS2B47NS3 Pro WT and its mutant L30S F31S.	183
Figure 4.13. Construct design and crystal structures of DENV4 NS3.	187
Figure 4.14. Comparison between overall conformations of NS2B47NS3 with previous structures.	188
Figure 4.15. In full length NS3 structures, the major crystal contacts are formed by the helicase domain.	189
Figure 4.16. The 2mFo-Fc maps of NS2B ₄₇ NS3 structures display solvent shell around the electron density of protein.	190
Figure 4.17. Different conformations of NS2B cofactor in full length NS3 structures.	194

Figure 4.18. Binding of BPTI to eNS2B ₄₇ NS3.	195
Figure 4.19. The domain arrangement between protease and helicase is brought by the loop 461-471 of helicase and linker region between the helicase and the protease domain.	196
Figure 4.20. Characterization of protease activity of NS3 full length constructs with different linkers.	198
Figure 4.21. The ATP hydrolysis activity of NS2B ₄₇ NS3 constructs.	200

List of tables

Table 1.1. List of cellular proteins involved in viral entry and binding	32
Table 1.2. Flavivirus polyprotein cleavage sites by NS2B-NS3 protease.	48
Table 1.3. Flavivirus vaccines currently under development	67
Table 1.4. List of flavivirus NS2BNS3 protease structures.....	85
Table 1.5. Peptidomimetic inhibitors of NS2B-NS3 protease.....	92
Table 1.6. Non-peptidic NS2B-NS3 protease inhibitors	95
Table 3.1. Data collection and refinement statistics for crystal structures of ZIKV protease (eZiPro and bZiPro).....	126
Table 3.2. Data collection and refinement statistics for bZiPro-fragment molecules structures	137
Table 3.3. Data collection and refinement statistics for bZiPro-dipeptide inhibitors structures	148
Table 3.4. Data collection and refinement statistics for bZiPro-P4-P1 peptidomimetic inhibitors	153
Table 4.1. Data collection and refinement statistics for eNS2B ₄₇ NS3 Pro.....	171
Table 4.2. Crystallization conditions and unit cell dimensions of NS2B ₄₇ NS3 constructs. ..	184
Table 4.3. Data collection and refinement statistics of NS2B ₄₇ NS3 full length from DENV	191
Table 6.1. Amino acid sequences of the constructs used in crystallography.....	209
Table 7.1. List of NS2B-NS3 structures reported in this thesis.....	211

Table of abbreviations

Abbreviations	Full Name
ADP	Adenosine diphosphate
AMP-PnP	Adenylyl-imidodiphosphate
AMC	Aminomethyl coumarin
ATP	Adenosine triphosphate
ATPase	Adenosine triphosphatase
Bz-nKRR-H	Benzoyl-norleucine-lysine-arginine-arginine-H
Bz-nKRR-AMC	Benzoyl-Nleu-Lys-Arg-Arg-Aminomethylcoumarin
β -ME	β -mercaptoethanol
CCP4	The Collaborative Computational Project number 4
CLEC5A	C-Type lectin domain containing 5A
CM	Convolute membrane
CS	Cyclization sequence (CS)
DC-SIGN	Dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin
ds	Double stranded
DF	Dengue Fever
DHF	Dengue haemorrhagic fever
DENV	Dengue virus
DTT	Dithiothreitol
DSS	Dengue shock syndrome
EEEV	Eastern equine encephalitis virus
ER	Endoplasmic reticulum
ETC	Experimental Therapeutics Centre
FL	Full length
GAG	Glycosaminoglycans
GTP	Guanosine triphosphate
GTPase	Guanosine triphosphatase
HCV	Hepatitis C virus
JEV	Japanese encephalitis virus
IPTG	Isopropyl- β -D-thiogalactopyranoside
IFN	Interferon

IFNAR	Interferon response alpha receptor
ISGs	Interferon stimulated genes
K_d	Dissociation constant
K_m	Apparent affinity
LAV	Live attenuated virus
LDLR	Low density lipoprotein receptor
LGP2	Laboratory of genetics and physiology 2
MALDI-TOF	Matrix-assisted-laser desorption ionization time-of-flight
MAD	Multiple wavelength anomalous dispersion
MES	2-morpholin-4-ethylsulfonic acid
MDA5	Melanoma differentiation associated protein 5
Mg	Magnesium
MHC-I	Major histocompatibility complex I
Mn	Manganese
MR	Molecular replacement
MTase	Methyl transferase
MVEV	Murray Valley encephalitis virus
NIAID	National Institute of Allergy and Infectious Diseases
NK	Natural killer
NLRs	NOD (nucleotide binding oligomerization domain)-like receptors
NS2B ₄₇	Residues 47-96 of NS2B protein
NS2B ₁₈	Residues 47-66 of NS2B protein
NS3	Non-structural protein 3
NS3 Hel	Non-structural protein 3 helicase domain
NS3 Pro	Non-structural protein 3 protease domain
NTP	Nucleotide triphosphate
NTPase	Nucleotide triphosphatase
PAMP	Pathogen associated molecular patterns
PEG	Polyethylene glycol
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PHEIC	Public Health Emergency of International Concern
P_i	Inorganic phosphate

PRRs	Pathogen-associated molecular pattern recognition receptors
RC	Replication complex
RIG-I	Retinoic acid inducible gene-I
RLRs	RIG-I like receptors
RdRp	RNA dependent RNA polymerase
RF	Double stranded replicative form
RI	Replicative intermediate
SAXS	Small angle X-ray scattering
SAM	S-adenosyl methionine
SAD	Single wavelength anomalous dispersion
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
ss	Single stranded
SL	Stem loop
STING	Stimulator of interferon genes
TAM	TYRO3, AXL, MER
TBEV	Tick-bone encephalitis virus
TIM	T-cell immunoglobulin and mucin domain
TLRs	Toll like receptors
Tris-HCl	Trisaminomethane-hydrochloric acid
UAR	Upstream of AUG region
UTR	Un-translated regions
WHO	World Health Organization
WNV	West Nile virus
ZIKV	Zika virus
YFV	Yellow fever virus

Summary

Flaviviruses are positive sense RNA viruses transmitted by arthropods. The flavivirus genus includes many important human pathogens such as Dengue Virus (DENV), West Nile virus (WNV), Yellow fever Virus (YFV), and recently emerged Zika Virus (ZIKV). About 40% of the world population is at risk of the flavivirus infections. The latest outbreaks of ZIKV in 2015 resulted in World Health Organization declaring Public Health Emergency of International Concern (PHEIC). There are neither specific nor broad-spectrum antivirals against flavivirus infections. The flavivirus polyprotein encodes three structural and seven non-structural proteins. Non-structural protein 3 (NS3) is the second largest protein encoded by the virus and composed of two domains, N-terminal protease and C-terminal RNA helicase/adenosine triphosphatase (ATPase) domains. The enzymatic activities of non-structural protein 3 are essential for viral RNA replication in the host cell making NS3 a prime target for the design of compounds with antiviral activity. In this study, 3D X-ray crystallography structures of NS2B-NS3 protease from ZIKV unbound and bound to different inhibitors are reported. The structures of ZIKV protease with different inhibitors compounds contributes significantly to the compound optimization in structure-based drug discovery. Biochemical assays were also conducted to characterize the NS2B-NS3 protease activity with and without inhibition by different compounds. In addition, structural and biochemical properties of the NS2B-NS3 protease and NS2B-NS3 full length from DENV, a closely related flavivirus, were studied. The structure of NS2B-NS3 protease from DENV reveals a new conformation which could be useful in virtual screening with compounds. We reported free enzyme structures and NS2B-NS3 complex in complex with bovine pancreatic trypsin inhibitor (BPTI). Altogether, the work in this thesis provides insights into the dynamics of the protease domain in full length context as well as conclusive biochemical activities of each domain from full length NS3 protein.

1. Chapter 1: Background

Chapter 1.1 History, taxonomy and Pathogenesis

Viruses were first discovered at the end of nineteenth century by three plant biologists, Mayer, Ivanowsky and Beijerinck during the study of tobacco mosaic disease (Gransden, 1999). They found that the infectious agent that caused the disease was not retained by the Chamberland-filters which ruled out bacteria as the causative agent (Leppard, 2007). Beijerinck demonstrated that the causative agent was incapable of independent growth and required the presence of host cells to replicate (Artenstein, 2012). By 1930s, more than 10 of these filterable agents had been shown to cause diseases such as yellow fever, rabies, pox, and foot-and-mouth disease (Leppard, 2007). The bacteriophage was discovered in 1910 and subsequently used in the famous Hershey-Chase experiment to prove that DNA was the genetic material and not the proteins (Hershey and Chase, 1952). As the discoveries of the viruses progressed, the International Committee on Nomenclature of Viruses was formed in 1966 (Leppard, 2007). In 1973, it was named as the International Committee on Taxonomy of Viruses (ICTV). The Baltimore classification invented in 1977 groups the viruses depending on the nature of their genetic materials (Leppard, 2007) (Figure 1.1).

Figure 1.1 Baltimore classification of viruses depending on their genetic materials. Adapted from (www.viralzone.expasy.org/254)

Section 1.1.1: *Flaviviridae* family viruses: overview on pathogenesis, molecular biology and anti-viral strategies

In 1960, a number of viruses responsible for causing encephalitis were grouped as group A arboviruses and group B arboviruses under the family, *Togaviridae* (Murphy et al., 2012). The group A arboviruses comprise eastern equine encephalitis, western equine encephalitis, Semikimi forest virus and group B arboviruses comprise Japanese encephalitis virus (JEV), West Nile virus (WNV), Yellow fever virus (YFV) and Murray valley encephalitis virus (MVEV) (Murphy et al., 2012). In 1984, based on the differences in structure, genetic sequence, and replication strategy, *flaviviruses* were classified as a separate family, *Flaviviridae*, while the Group A arboviruses becomes the genus, *Alphaviruses* (Brinton, 1986; Holbrook, 2017a).

Flaviviridae are a family of positive sense RNA viruses. It consists of four genera, *Flavivirus*, *Pestivirus*, *Hepacivirus* and *Pegivirus* (Simmonds et al., 2017) (Figure 1.2). The *Pestivirus* and *Hepacivirus* do not have arthropod-borne vector. The *Pestivirus* infects cattle and includes bovine viral diarrhoea virus and classical swine fever virus (Simmonds et al., 2017). The best characterized member of the genus *Hepacivirus* is hepatitis C virus (HCV). Its human infections lead to chronic hepatitis and liver cirrhosis (Simmonds et al., 2017). *Pegivirus* causes persistent infections in a range of mammalian hosts and are much less studied (Simmonds et al., 2017; Stapleton et al., 2011). The genus *flavivirus* is transmitted by tick or mosquito mostly. Many human pathogens are mosquito borne flaviviruses, eg. DENV, ZIKV, WNV and YFV, etc. (Simmonds et al., 2017). The genome organizations of the different genera of *Flaviviridae* are shown in Figure 1.3.

Figure.1.2 Phylogenetic tree based on the NS5 RNA dependent RNA polymerase (RdRp) of *Flaviviridae* family. The lengths of the lines are directly proportional to genetic distance. The tree is constructed using bootstrap method. This figure is adapted from (Schweizer and Peterhans, 2014)

Pestivirus

Pestiviruses are a member of *Flaviviridae* family with species such as bovine viral diarrhoea virus (BVDV), classical swine fever virus (CSFV) that infect cattle and livestock animals (Schweizer and Peterhans, 2014). Similar to other members of the *Flaviviridae* family, these viruses have +ssRNA genome associated with capsid protein surrounded by lipid membrane and envelope glycoprotein (E proteins) (Schweizer and Peterhans, 2014; Tautz et al., 2015). In contrast to other flaviviruses, these viruses can exist in two different biotypes, non-

cytopathic and cytopathic biotypes (Peterhans et al., 2010; Peterhans and Schweizer, 2010; Schweizer and Peterhans, 2014; Tautz et al., 2015). While the cytopathic strains are responsible for causing disease in livestock, only non-cytopathic strains can establish persistent infections (Schweizer and Peterhans, 2014). The virus enters the cells by E protein mediated attachment, followed by clathrin mediated endocytosis, and low pH induced fusion and release of viral genome into the cytoplasm (Grummer et al., 2004; Krey et al., 2005). The viral replication takes place on ER membrane. The virus maturation takes place in intracellular vesicles and the newly synthesized viruses leave the cell via exocytosis (Tautz et al., 2015). The genome is 12.3 kb long, flanked by 5' and 3' UTRs and encodes a non-structural protein called N^{pro}, capsid protein, three E proteins, and seven non-structural proteins, p7, NS2, NS3, NS4A, NS4B, NS5A and NS5B (Dehan et al., 2001; Tautz et al., 2015). The 5' end of the genome is capped and contains internal ribosomal entry site (IRES) (Tautz et al., 2015). The N-terminal protease (N^{pro}) is a cysteine protease and cleaves its own C-terminus while capsid protein is cleaved by cellular signal peptidase (Neill, 2013; Schweizer and Peterhans, 2014; Tautz et al., 2015). The signal peptidase is also responsible for cleavage at Erns, E1, E2, p7 and NS2 boundaries. The NS2, NS3, NS4B, NS5A and NS5B cleavages are carried out by NS3 protease together with NS4A cofactor (Tautz et al., 2015). Transient infections by cytopathic strain show symptoms of viremia. Infection with non-cytopathic or cytopathic strain during the pregnancy could lead to persistent infections in new-born calves where they become immunotolerant to the virus and prone to mucosal disease (Brownlie, 1990; Gallei et al., 2008; Schweizer and Peterhans, 2014). Due to its importance and economic impact on livestock, vaccines for BVDV have been developed over the years. Currently modified-live viral and inactivated viral vaccines are available against BVDV and CSFV.

Hepacivirus

The most studied species of genus *Hepacivirus* is hepatitis C virus (HCV). HCV is responsible for most of viral induced chronic hepatitis, cirrhosis and liver cancer. HCV was discovered in 1989 by cloning the cDNA from the infectious agent which caused non-A, non-B hepatitis. HCV causes persistent infections in liver that are difficult to detect although the virus replicates very well in the organ (Lindenbach and Rice, 2003a). HCV mostly targets hepatocytes but is also able to infect other cell types such as B cells and dendritic cells. A number of cellular receptors such as CD81, scavengers receptor class B type I (SR-B1), C-type lectin and low density lipoprotein receptor (LDLR) are responsible for HCV attachment and entry but none of them alone is sufficient for viral attachment and entry (Moradpour et al., 2007). Upon binding to the cell surface, the virus is taken up into the cell via clathrin coated endosome where the low pH endosomal environment induces the viral membrane fusion leading to the release of viral genome (Lindenbach and Rice, 2003b; Moradpour et al., 2007). In the cytoplasm, HCV replicates on the ER membrane (Paul et al., 2014). The viral replication remodels the ER membrane to form double-membrane vesicles (DMV) where the replication complex localizes to (Moradpour et al., 2007; Paul et al., 2014). The genome of HCV is 9.6 kb long and consists of 5' non-coding regions (NCR), followed by three structural proteins, core (C) protein and two envelope proteins, E1 and E2. Downstream of the structural proteins are seven non-structural proteins, p7, NS2, NS3, NS4A, NS4B, NS5A and NS5B (Lindenbach and Rice, 2005). Viral replication is initiated by IRES (Ji et al., 2004; Spahn et al., 2001). Polyprotein processing is performed by both cellular and viral proteases (Lindenbach and Rice, 2005). The cleavages between C, E1, E2 and p7 are carried out by host signal peptidase (Lin et al., 1994). NS2 cysteine autoprotease processes its own NS2B/3 junction while the rest of the junctions between non-structural proteins are cleaved by NS3

serine protease together with NS4A cofactor (Apte-Sengupta et al., 2014; Failla et al., 1994; Hahm et al., 1995; Han et al., 1995).

Pegivirus

Pegivirus contains species that cause hepatitis in human and primates (Stapleton et al., 2011). The most characterized pegivirus is the hepatitis G virus (HGV) also called GBV-C virus (Stapleton et al., 2011). GBV-C infection is highly common in humans with 15% of blood donors having previous or ongoing infections (Knipe, 2007). Similar to HCV, GBV-C also causes persistent infections (Stapleton et al., 2011). The viral replication cycle of GBV-C is also similar to HCV although GBV-C has very long 5' UTR distinct from the rest of the genera in the family (Knipe, 2007; Stapleton et al., 2011). The genome replication is via IRES with cap independent initiation mechanism (Simons et al., 1996; Yoo et al., 1992). Interestingly, GBV-C does not contain capsid protein; hence it is assumed that virus utilizes host capsid like protein (Muerhoff et al., 1995; Theodore and Lemon, 1997; Xiang et al., 1998). GBV-C contains E1 and E2 glycoproteins as well and shows similar non-structural protein boundaries, such as NS2, NS3, NS4A, NS4B, NS5A and NS5B. GBV-C encodes p13 which is similar to p7 of HCV (Takikawa et al., 2006). The protease activity of NS3-4A and *de novo* RdRp initiation and elongation activities of NS5B have been demonstrated (Leary et al., 1996; Sbardellati et al., 2000; Xiang et al., 2008).

Flavivirus

The genus *Flavivirus* contains many human pathogens, such as West Nile virus (WNV), Yellow fever Virus (YFV), Dengue Virus (DENV) and Zika Virus (ZIKV). The flavivirus genome is 11 kb long with 5' cap. The genome consists of three structural proteins and seven non-structural proteins. The genome is translated as one polyprotein with one open reading frame (ORF). The host peptidase and viral peptidase further process the polyprotein into

individual structural and non-structural proteins. The details on molecular virology of flavivirus are discussed in Chapter 1.2 – 1.4.

Figure 1.3. Genome organization of different genera of *Flaviviridae*. (A) Pestivirus (B) Hepacivirus (C) Pegivirus (D) Flavivirus.

Flavivirus epidemiology, phylogeny and pathogenesis

The genus flavivirus of *Flaviviridae* family contains a number of human pathogens. The *Flaviviruses* can be grouped into tick borne, mosquito borne and no-known vector flavivirus (Figure 1.4). They can be separated into two groups depending on the type of vector used for transmission, tick or mosquito. The mosquito-borne flavivirus are serologically more diverse than the tick-borne ones (Holbrook, 2017a). Many human pathogens such as Dengue virus (DENV), Japanese encephalitis virus (JEV), West Nile virus (WNV) and Zika virus (ZIKV) are mosquito borne flaviviruses (Figure 1.4) (Kuno et al., 1998; Simmonds et al., 2017). The geographical distribution of flaviviruses that are important human pathogens are shown in Figure 1.5. Currently, about 40% of the world population is at risk of infection. With global warming and urbanization, the flaviviruses impose a serious public health threat.

Figure 1.4. Phylogenetic tree of members of *flavivirus* genus based on polyprotein precursor.

Figure adapted from (Ashraf et al., 2015)

Figure.1.5. Geographical distribution of flaviviruses that are important human pathogens.

Figure adopted from www.creative-diagnostics.com/Flavivirus.htm.

The mosquito borne flaviviruses could be separated into two groups depending on the type of the mosquito vector. The flaviviruses that are transmitted by *Culex* spp mosquitoes are generally associated with neurotropic diseases, while those transmitted by *Aedes aegypti* mosquitoes are more closely associated with viscerotropic diseases (Holbrook, 2017a). The transmission cycles of the flaviviruses can be separated into three, (1) urban epidemic cycle (2) enzootic cycle and (3) epizootic cycle (Figure 1.6) (Vasilakis and Weaver, 2017; Weaver

and Barrett, 2004). The flaviviruses that use *Culex spp* as the vector are transmitted via enzootic cycle or epizootic (eg. WNV and JEV)(Colpitts et al., 2012). In an enzootic cycle, the disease caused by the virus is prevalent in the non-human animals in a specific region or during a certain climate in an expected manner (Go et al., 2014). In an epizootic cycle, the disease outbreak occurs in higher frequency or spread than the expected occurrence (Go et al., 2014). The viruses whose vector is *Aedes* mosquito are transmitted either by enzootic cycle or urban epidemic cycle (Vasilakis and Weaver, 2017; Weaver and Barrett, 2004; Zamani and Zamani, 2017). In enzootic and epizootic cycle, humans are dead end hosts due to their low level viremia that is not sufficient to transmit to mosquito, ending the transmission cycle (Colpitts et al., 2012). Should the JEV and WNV adapt to *Aedes* mosquito, it is possible that these viruses will also switch to urban transmission cycle (Colpitts et al., 2012).

Upon feeding on infected blood, the virus is taken up by the female mosquito. To be transmitted by the mosquito, the virus must first replicate in the midgut followed by the release into the hemolymphs (Salazar et al., 2007; Sriurairatna and Bhamarapavati, 1977). The virus further replicates in the hemolymph and reaches the salivary glands. The third replication occurs in salivary glands from which the virus is transmitted with the next bite. The efficiency of transmission by different types of mosquitoes depends both on the type of mosquito and the virus species. One of the common barriers for the efficient transmission is due to the virus inability to cross the mesenteron. *Culex* mosquitoes infected with ZIKV were unable to transmit the virus. Although the virus was detected in the midgut of the mosquito, it was not detected in the salivary glands and head of the mosquitoes indicating that the incompetence of *Culex* mosquitoes for ZIKV infections are possibly due to the midgut barrier (Liu et al., 2017). However, it is to be noted that there are studies claiming that *Culex* mosquitoes can transmit ZIKV (Ayres et al., 2017; Guedes et al., 2017; Roundy et al., 2017). In another study, it was shown that *Anopheles* mosquitoes were transmission competent via

intrathoracic inoculation with Murray Valley encephalitis virus (MVEV) but not via ingestion of infected blood meal (Mc, 1955). Merrill and Tenbroeck also found that *A. aegypti* could transmit eastern equine encephalitis virus (EEEV) only after the abdomen was punctured by a needle following an infected blood meal (Merrill and Tenbroeck, 1935).

Figure 1.6. Transmission cycles of flaviviruses. (Vasilakis and Weaver, 2017). In the urban epidemic cycle, the virus circulates between humans and domestic/peridomestic *Aedes* mosquitoes. In the enzootic and epizootic cycle, the virus circulates between the mosquito and non-human vertebrates with humans being dead-end hosts. During the enzootic cycle, the direct spill over to humans occurs when humans come into contact with the sylvatic habitats. In epizootic cycle, due to prolonged exposure of human habitats to the sylvatic population, the virus has evolved to infect domestic animals and has sustained a transmission cycle between the domestic animals and the mosquitoes with spill overs to human hosts (Weaver and Barrett, 2004; Weaver et al., 2016).

One of the well documented flavivirus is YFV, the representative virus of the genus *Flavivirus*. The first outbreak of yellow fever was documented in 1647 in Barbados (Holbrook, 2017a). Subsequently, the YFV reached the Caribbean and to South Americas followed by North Americas (Jentes et al., 2011) (Figure 1.5). Although clinical manifestations, disease symptoms and outbreaks were documented as early as the eighteenth century, the causative agent, yellow fever virus was discovered only in 1928 (Stokes, 1928). The infection with YFV could be sub-clinical mild form or more severe biphasic haemorrhagic form (Monath and Barrett, 2003). The biphasic stage of yellow fever consists of the mild “infection” phase, with a brief period of remission, followed by the severe “intoxication” phase (Monath and Barrett, 2003). The severe haemorrhagic fever and acute liver injury happen during the “intoxication” phase (Monath and Barrett, 2003). YFV directly infects the hepatocytes and Kupffer cells causing acute liver dysfunction characterized by jaundice (Holbrook, 2017a). The first yellow fever vaccine was reported in 1937 by Max Theiler (Theiler and Smith, 1937). The vaccine contains attenuated YFV achieved by serial passage of the virus in mouse. Since then YFV 17D vaccine has minimized the potential YFV outbreaks including a recent YFV outbreak in Angola in 2016 (Holbrook, 2017a; Ishikawa et al., 2014).

Another important human pathogen in the genus *Flavivirus* is DENV, the causative agent of dengue fever (Holliday, 1880). With pandemics occurring about every 50 years from 18th century till present day, dengue imposes a major public health burden (Halstead et al., 2007). Similar to YFV, Dengue manifests in two forms, sub-clinical flu like symptoms representing the mild form, and severe forms represented by dengue haemorrhagic fever (DHF) and dengue shock syndrome (DSS) (Halstead, 2007). The DENV has four distinct serotypes. The severity of dengue has been linked to infection with more than one serotype (Wan et al., 2013). The primary infection normally results in mild dengue fever. However, due to the

presence of cross-reactive, non-neutralising antibodies after primary infection, subsequent reaction with a different serotype results in higher number of infected cells, leading to a more severe form of dengue (Flipse et al., 2013; Halstead, 2014). This process is known as antibody dependent enhancement (ADE) (Halstead, 2014). The development of DENV vaccine has been hampered by the need for the vaccine to be cross-protective against all serotypes of DENV without inducing the ADE. Although the first approved DENV vaccine was available in 2016, recent reports of the severe DENV form linked to the vaccinated individuals who were previously not infected by DENV have raised questions on the efficacy of the vaccine (Aguilar et al., 2017; Aguilar et al., 2016; Villar et al., 2015; Yang et al., 2018).

Japanese Encephalitis virus is another flavivirus which affects about 70,000 people a year with mortality of 14,000-20,500 per year and is transmitted by *Culex* mosquitoes (Campbell et al., 2011). The transmission of JEV is via enzootic cycle involving pigs acting as an amplifying host (Weaver and Barrett, 2004). Symptoms of JEV infection include headache, diarrhea and vomiting during the mild stage and can cause more severe neuropathological symptoms such as acute flaccid paralysis, cogwheel rigidity, meningitis, coma and death (Hills and Phillips, 2009; Sohn, 2000). About 30% of the recovered patients showed neuropathology such as seizures and cognitive disabilities. The first formalin-inactivated vaccine was introduced in Japan in 1954 (Yun and Lee, 2014). Since then several other forms of attenuated or inactivated JEV vaccines have become available in Asian markets (Yun and Lee, 2014). A chimeric vaccine with prM and E proteins from JEV fused to YFV 17D was introduced in Thailand and Australia by Sanofi Pasteur (Yun and Lee, 2014).

Tick borne encephalitis virus was discovered in early 20th century in far eastern Soviet Union (Holbrook, 2017a). As the name implied, TBEV is transmitted by *Ix ricinus* tick and can cause mild to severe form of encephalitis with the potential of causing long term neuropathological effects such as paresis and atrophy of neck muscles, lower extremities

muscles similar to poliomyelitis infection(Holbrook, 2017a). The first TBEV vaccine was developed in 1937 (Ishikawa et al., 2014). Since then, various inactivated strains of TBEV have been used to produce vaccines in Europe (Ishikawa et al., 2014; Lehrer and Holbrook, 2011).

The latest emerging flavivirus is Zika virus (ZIKV). Despite its early discovery in 1947 in Uganda, ZIKV only came into the spot light during the re-emergences in recent years after its infections were linked to new born microcephaly and adult neuropathy such as Guillain-Barré syndrome (Figure 1.5) (Brasil et al., 2016; Cao-Lormeau et al., 2016; Dick et al., 1952; Faria et al., 2016). For 60 years since its discovery, no ZIKV epidemic was reported partly due to its symptoms being similar to dengue and other flaviviral infections. The first ZIKV epidemic was reported in 2007 in Yap islands, Micronesia. The second ZIKV outbreak was in French Polynesia, in 2013 with 30,000 cases along with neurological complications (Gubler et al., 2017). Most ZIKV infections are similar to DENV with fever, rash, and muscle pain. However, in some cases ZIKV were linked to adult and infant neuropathy (Cao-Lormeau et al., 2016; Mlakar et al., 2016; Rasmussen et al., 2016; Vesnaver et al., 2017). In addition to horizontal host-vector transmissions like other flaviviruses, ZIKV could also be transmitted from sexual intercourse and from infected mother to fetus (Brasil et al., 2016; D'Ortenzio et al., 2016; Frieden et al., 2016; Musso et al., 2015). Currently there is no antiviral or vaccine available for ZIKV.

Chapter 1.2: Molecular biology of Flavivirus

Section 1.2.1: Flavivirus life cycle

Flaviviruses are positive sense single stranded RNA viruses. Flavivirus enters the cell by attachment to the cell surface via glycoproteins, interaction with cell-surface receptors followed by endocytosis (Mukhopadhyay et al., 2005). In the case of DENV infections, the

presence of cross-reacting, non-neutralizing antibodies due to prior infection with a different DENV serotype will lead to higher uptake of virus into the cells leading to ADE (Halstead, 2014; Holbrook, 2017b; Weaver et al., 2016; Wen and Shrestha, 2017). Due to ADE, subsequent infections with different serotype of DENV are associated with higher risk of developing into severe DENV. Previous infections with other flavivirus have been shown to enhance the efficiency of infections in ZIKV as well (Bardina et al., 2017; Wen and Shrestha, 2017). Upon entering the cell, the acidic pH in the endosome triggers conformational changes of E protein which in turn induce fusion of the viral envelope with endosomal membrane leading to release of the viral RNA genome (Hasan et al., 2018). The +ssRNA genome is translated to viral polyproteins in the ER lumen starting the viral replication process (Lindenbach and Rice, 2003b). Flavivirus genome is 11kb long flanked with 3' and 5' untranslated regions (UTR) (Apte-Sengupta et al., 2014). Genome encodes three structural proteins and seven non-structural proteins as one polypeptide which is processed by viral and host proteases into individual proteins (Lindenbach and Rice, 2003b). Replication takes place on the ER membrane in specialized membrane compartments induced by the non-structural proteins and their interaction with host proteins (Paul and Bartenschlager, 2015). The flavivirus life cycle is shown in Figure 1.7. The virus encodes its own RNA dependent RNA polymerase (RdRp) (Chambers et al., 1990a). The major enzymatic activities in the replication complex are carried out by non-structural protein 3 and non-structural protein 5 (Lindenbach and Rice, 2003b). Non-structural protein 3 is a multifunction domain with N-terminal protease and C-terminal helicase/nucleotide triphosphatase (NTPase) (Chambers et al., 1991; Chambers et al., 1993; Falgout et al., 1991; Li et al., 1999; Lindenbach and Rice, 2003b). Non-structural protein 5 is a RNA dependent RNA polymerase with methyltransferase (Ackermann and Padmanabhan, 2001; Lindenbach and Rice, 2003b). The genome assembly and packaging of immature virus take place in the ER (Lindenbach and

Rice, 2003b; Mukhopadhyay et al., 2005). The immature viral particles are transported to the trans-golgi network where prM is cleaved by furin(Lindenbach and Rice, 2003b). The mature viral particle is released by exocytosis (Lindenbach and Rice, 2003b).

Figure 1.7. Flavivirus replication cycle. Figure adapted from (Gerold et al., 2017)

Virus binding and entry

Receptors responsible for binding and entry

The flaviviruses can infect different cell types in a range of host species indicating that the receptors utilized by the virus for binding and entry could be ubiquitous and coordinated involving more than one binding receptor (Perera-Lecoin et al., 2014; Rey et al., 2017). In

recent years, a number of cell surface receptors have been identified as flavivirus entry receptors in different cell lines with some receptors being common between the species while others being species specific (Perera-Lecoin et al., 2014; Sirohi and Kuhn, 2017). The interactions between E protein and cell surface receptors facilitate virus attachment to the cells (Perera-Lecoin et al., 2014; Rey et al., 2017). One of the earliest receptors discovered involved in viral attachment were negatively charged glycosaminoglycans (GAG) such as heparan-sulfate (Chen et al., 1997b). These abundant and ubiquitously expressed GAGs are common receptors for different flaviviruses (Perera-Lecoin et al., 2014). These GAGs bind to the E protein domain III via non-specific electrostatic interactions and concentrate the viral particles on the cell surface (Perera-Lecoin et al., 2014; Rey et al., 2017; Sirohi and Kuhn, 2017). C type lectins such as dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin (DC-SIGN) and mannose receptors are transmembrane cell surface receptors expressed on dendritic cells and macrophages. DC-SIGN binds to the two glycosylation sites on E protein. Single polymorphism in DC-SIGN has been associated with severe form of DENV and TBEV (Barkhash et al., 2012; Wang et al., 2011). The DC-SIGN has also been reported as a receptor for ZIKV (Hamel et al., 2015). The mannose receptor could bind to all four serotypes of DENV (Tassaneetrithep et al., 2003). The mannose receptor is expressed in macrophages and involved in uptake of antigen presenting cells via clathrin-dependent endocytosis. C-Type lectin domain containing 5A (CLEC5A) is expressed in monocytes and macrophages (Lok et al., 2008). Although CLEC5A interacts with E protein it is not yet proven that the glycans on E proteins are responsible for the interaction. Binding of virus to CLEC5A induces the release of pro-inflammatory cytokines and could be responsible for severe form of flaviviral infections (Chen et al., 2012). Another important class of receptor group is T-cell immunoglobulin and mucin domain (TIM) receptors and TYRO3, AXL, MER (TAM) receptors (Jemielity et al., 2013). Both TIM and TAM receptors

interact with apoptotic cells and promote the endocytosis of apoptotic bodies via clathrin-coated endosome. TIM ectoexpression can enhance many flaviviral infections such as DENV, WNV and YFV. Silencing of TIM receptors on permissive cell lines has inhibited the WNV and DENV infections. TAM binding of apoptotic cells occurs via bridging molecule Gas6/ProS and it has been demonstrated that it is the same for flaviviruses as well. The TAMs were identified as receptors responsible for ZIKV infection in human skin cells along with DC-SIGN but interestingly, role of TIM in ZIKV infections were not significant (Hamel et al., 2015).

Table 1.1. List of cellular proteins involved in viral entry and binding

Molecule	Properties	Cell line	Virus	References
Heparan sulfate	Glycosaminoglycans	Mammalian	DENV, JEV, MVEV, TBEV, YFV, WNV, ZIKV	(Chen et al., 1997b; Germi et al., 2002; Kim et al., 2017; Kroschewski et al., 2003; Lee et al., 2006)
HSP70/90	Heat-shock protein	Mammalian, mosquito	DENV, JEV	(Das et al., 2009; Reyes-Del Valle et al., 2005)
DC-SIGN/L-SIGN Mannose receptor CLEC5A mosGCTL-1	C-type lectins	Mammalian Mammalian Mammalian Mosquito	DENV, WNV, ZIKV DENV DENV, JEV WNV,	(Chen et al., 2008; Chen et al., 2012; Cheng et al., 2010; Davis et al., 2006; Dejnirattisai et al., 2011; Hamel et al., 2015; Tassaneetrithep et al., 2003)
Laminin receptor	Laminin receptor	Mammalian, mosquito	DENV, WNV	(Sakoonwatanyoo et al., 2006; Thepparit and Smith, 2004)
TIM receptor TAM receptor	Phosphatidylserine receptors	Mammalian	DENV, WNV, YFV, DENV, WNV, YFV, ZIKV	(Jemielity et al., 2013; Kuadkitkan et al., 2010) (Meertens et al., 2017)
Scavenger receptor Class B type I	High density lipoprotein receptor	Mammalian	DENV	(Li et al., 2013)
Claudin-I	Tight junctions	Mammalian	DENV	(Che et al., 2013; Gao et al., 2010)

Viral fusion

After binding to the entry receptors / attachment factors, the virus enters the cell via clathrin coated pits where the pits form invagination in the plasma membrane (Chu and Ng, 2004). Treatment of cells with inhibitors that prevent the formation of clathrin coated pits results in inhibition of WNV entry (Chu et al., 2006). However the clathrin independent entry has been observed for DENV infection in mammalian cell lines (Acosta et al., 2009). Following clathrin mediated endocytosis, the flaviviral particles localise in early endosomes which in turn mature into late endosomes, followed by membrane fusion (Chao et al., 2014). Viral particle internalization and endosomal fusion take place about 10-15 minutes after infection (van der Schaar et al., 2007). The envelope protein (E protein) of flaviviruses is made up of transmembrane domain and ectodomains. The ectodomains of E protein are domain I (DI), domain II (DII) and domain III (DIII) (Figure 1.8A)(Kanai et al., 2006; Kaufmann et al., 2009; Li et al., 2008a; Li et al., 2008b). During the pre fusion state, the E proteins are in dimer state (Figure 1.8A) (Chao et al., 2014; Chu and Ng, 2004; Kanai et al., 2006; Kaufmann et al., 2009; Yu et al., 2008). The membrane fusion is triggered by the low pH environment in the late endosome while studies have suggested pro-fusion of cholesterol and anionic lipids (Yu et al., 2008; Zhang et al., 2003). Mildly acidic pH in endosomal compartments disturbs the E protein dimers via the protonation of residues resulting in E protein monomers (Li et al., 2008a; Rey et al., 1995a) (Chao et al., 2014; Nybakken et al., 2006; Zhang et al., 2015). The DII of E protein monomers insert the now exposed fusion loop peptide into the endosomal membrane (Figure 1.8 B) (Yu et al., 2008; Zhang et al., 2004). The outwardly projected E monomers form trimers via DI domains (Yu et al., 2008). These trimers are not stable resulting in DIII folding back like a hairpin bringing the endosomal

membrane close to the viral membrane resulting in the merging of the membranes, and subsequently forming a fusion pore and therefore releasing the nucleocapsid into the cytoplasm (Figure 1.8 E-F) (Modis et al., 2004).

Figure 1.8. Mechanisms of viral fusion with the envelope. (A) Pre fusion state (B) Changes in pH result in fusion loop which is inserted into the endosomal membrane. (C) E dimers are rearranged to monomers followed by trimer formations (D). (E) The trimers are unstable and fold back pulling the two membranes close. (F) Post fusion state where two membranes are joined releasing the viral genome. Figure adapted from (Chao et al., 2014)

Genome synthesis and replication

Flavivirus genomic RNA is positive stranded with a 5' cap and hence the genomic RNA can recruit the cellular translation machinery for translation of RNA into viral proteins (Lindenbach and Rice, 2003b). The viral genome replication and RNA synthesis occurs in single membrane vesicles on the ER membrane (Lindenbach and Rice, 2003b). The replication complex is located at the membrane vesicles (Paul and Bartenschlager, 2015). The vesicles can have small openings at the neck suggesting the exchange of nucleotides and RNAs with the cytoplasm (Paul and Bartenschlager, 2015; Welsch et al., 2009). Membrane rearrangements on the ER are brought about by viral and cellular proteins as well as the lipid composition (Apte-Sengupta et al., 2014; Perera et al., 2012). The non-structural membrane proteins NS2B, NS4A, NS4B are involved in membrane rearrangements (Apte-Sengupta et

al., 2014; Hasan et al., 2018). Interactions between NS3 and NS5 and other non-structural proteins have been observed indicating the role of NS2B, NS4A and NS4B in anchoring the replication complex to membrane (Gerold et al., 2017). During RNA synthesis, three RNA forms are observed, (1) double stranded replicative form (RF) RNA, (2) single-stranded genome length RNA and (3) replicative intermediate where nascent RNAs are associated with – strand template (RI)(Apte-Sengupta et al., 2014; Ng et al., 2017). The RNA synthesis is not symmetrical with genomic RNA accumulating more than the negative strand. The 5' cap is only present on genomic RNA and not on dsRNA intermediate. The 5' capping of genomic RNA involves 3 steps, (1) NS3 helicase hydrolyzes the 5' triphosphate end of RNA to 5' diphosphate, (2) guanosine triphosphate (GTP) binds to NS5 methyltransferase (MTase) and hydrolyzes to GMP and is subsequently transferred to 5' end of RNA. (3) NS5 MTase methylates the capped RNA at N7 of the guanine cap and at 2'-OH of the first nucleotide (Klema et al., 2015; Zhao et al., 2015a; Zhao et al., 2015b). The RdRp synthesizes the RNA via two modes, *de novo* initiation followed by elongation (Klema et al., 2015). The MTase domain might be involved in coordination of initiation and elongation activities of NS5 (Dong et al., 2010)..

Virion packaging and maturation

The flavivirus genome is packaged with capsid proteins to protect it from degradation by cellular ribonuclease and detection by immune system (Lindenbach and Rice, 2003b; Perera and Kuhn, 2008). The interaction between capsid protein and ssRNA is not specific yet the nucleocapsid is packaged into virions demonstrating that the genome replication and assembly are coupled (Hasan et al., 2018; Lindenbach, 2013; Perera and Kuhn, 2008). Nonetheless, the prM/E glycoproteins do not interact with nucleocapsid and subviral particles could be obtained by overexpression of prM/E proteins (Konishi and Mason, 1993). The sequential cleavage of C by NS2B/NS3 protease followed by the cleavage between C/prM

has been suggested as temporal regulation to make sure that mature capsid is released first before the prM is released (Bollati et al., 2010; Perera and Kuhn, 2008). Budding virions have been observed near the opening of vesicles packets where the genome RNA synthesis takes place (Gillespie et al., 2010; Miorin et al., 2013; Paul and Bartenschlager, 2015; Welsch et al., 2009). However, it is still not clear how the viral genomic RNA is transferred to the site of assembly (Lindenbach, 2013).

The newly synthesized immature virions bud into the ER lumen (Lindenbach and Rice, 2003b). The immature virions at this stage have a spiky appearance with prM located at the end of the spikes shielding the fusion loop (Kuhn et al., 2002b; Sirohi and Kuhn, 2017; Zhang et al., 2003; Zhang et al., 2004). As the virions are transported along the trans golgi network (TGN), the change in pH results in rearrangement of prM/E heterodimers (Li et al., 2008a; Mukhopadhyay et al., 2005; Sirohi and Kuhn, 2017)}. The prM cleavage site is now exposed to and recognized by resident protease furin which cleaves off the pr portion (Li et al., 2008a; Yu et al., 2008). Exocytosis of the virions to the extracellular neutral pH results in shedding of pr portion resulting in smooth surface mature virions (Yu et al., 2008). The prM cleavage is not an efficient process with partially mature virions being released into the extracellular space (Junjhon et al., 2010). Although partially immature virions are not as infectious as mature virions, it could be one of the immune evasion strategies by virus to saturate the antibodies (Flipse et al., 2013; Junjhon et al., 2010; Zhang et al., 2003). The prM antibodies in the serum of infected patients have been observed and have been linked to disease severity, indicating that partially mature virus has a role in pathogenesis (Rodenhuis-Zybert et al., 2011; Sirohi and Kuhn, 2017).

The virion structure

The first flaviviral structure was reported in 2002 by Kuhn et al., (Kuhn et al., 2002b). Since then various studies have reported different stages of flaviviral particles along with antibodies

and glycoproteins (Kuhn et al., 2002b; Li et al., 2008a; Lok et al., 2008; Zhang et al., 2003). The mature flavivirion is small and about 50 nm in diameter (Figure 1.9) (Lok et al., 2008; Yu et al., 2008; Zhang et al., 2003). The virus surface consists of membrane protein (M protein) and envelope protein (E protein). One virion is made up of 180 copies of E and M proteins each (Kuhn et al., 2002b; Mukhopadhyay et al., 2005; Perera et al., 2008). At the core of the virion, the +ssRNA genome is packaged with capsid protein and surrounded by host derived lipids. The structural information of nucleocapsid in the core is not available currently (Mukhopadhyay et al., 2005). However, the crystal structure of WNV C-protein shows a tetramer with positively charged 11 Å channel which can accommodate ssRNA (Dokland et al., 2004a). In mature virus, the E protein is arranged in dimers. The E protein is about 53 kDa and is anchored to the lipids by two transmembrane helices (Figure 1.9). The ectodomain part of E protein consists of three proteins, amino terminal domain (DI), dimerization domain (DII) and carboxyl terminal Ig like domain (DIII). E protein is responsible for binding and fusion of viral envelope with endosomal membrane during the entry and is the major antigenic component of the virus (Lindenbach, 2013). E protein dimers are arranged in trimeric rafts and forms viral particles with a smooth surface (Zhang et al., 2004). The immature viral particles have a spiky appearance compared to matured virus and are slightly larger about 60 nm in size (Figure 1.9) (Kuhn et al., 2002b). The surface of the immature virion is made up of 60 E proteins trimers which are associated with the prM (Zhang et al., 2015). The association with prM shields the E-protein fusion peptide from premature fusion during the transport across the trans golgi network (TGN) (Figure 1.9). During the maturation process, the pr- portion is cleaved by furin protease and E-protein rearrange to form homodimer and mature virus is released. After entering the cell, low pH environment of endosome causes conformational changes in E-protein to expose the fusion peptide (Figure 1.9) (Rey et al., 2017; Yu et al., 2008). The E-protein dimers dissociate into

monomers which then rearrange itself into trimers with fusion peptides on each monomer exposed towards the endosomal membrane to initiate the fusion (Rey et al., 2017). This intermediate virion structure is termed as fusogenic form (Zhang et al., 2015).

Figure 1.9. Mature and immature viral particles of DENV. The figure is adapted from (Rey et al., 2018) and (Hasan et al., 2018)

Genome organization and UTRs

Flavivirus has a single stranded positive RNA genome of about 11 kb long. The genome has 5' type I cap with m⁷GpppAmp but lacks the polyadenylated tail (Figure 1.10A) (Gebhard et al., 2011; Ng et al., 2017). The genome encodes a polyprotein with one open reading frame (ORF) and is flanked by untranslated regions of about 100 nucleotides (nts) at the 5' end and about 300-700 nucleotides at the 3' (Figure 1.1A and B). The 5' UTR consists of two secondary structures, stem loop A and stem loop B (SLA and SLB) Figure 1.10B (Ng et al., 2017). The 5' UTRs are not sequence conserved among the flavivirus but their secondary structures are conserved. The SLA acts as the promoter to recruit NS5 for RNA synthesis (Gebhard et al., 2011; Ng et al., 2017). The SLB region includes 5' upstream of AUG sequence (UAR) important for viral genome cyclization (Ng et al., 2017). The 3' UTR is longer and the sequence length varies within the species of flavivirus (Ng et al., 2017). Immediately after the STOP codon, 3' UTR has a variable region with different nucleotide length depending on the virus species (Gebhard et al., 2011). The host exonuclease XRN1 processes the viral genome in this region to produce sfRNA (Ng et al., 2017). This variable region is followed by two relatively conserved dumbbell structures that are involved in genome replication, translation and sfRNA synthesis (Gebhard et al., 2011; Lodeiro et al., 2009). The last stretch of 3' UTR is the most conserved and it contains 3' cyclization sequence (CS), 3' upstream of AUG region (UAR) and 3' stem loop (SL). The 3' CS and UAR are involved in genome cyclization while the 3' SL interacts with host and viral protein and regulates between translation and RNA synthesis (Ng et al., 2017).

Figure 1.10. Genome organization of flavivirus. The secondary structures at the untranslated regions of 5' and 3'. Figure adapted from (Ng et al., 2017) and Viralzone.

Section 1.2.2: Structural proteins from flavivirus

Capsid protein

The capsid protein is a small protein of about 11 kDa with N and C-terminal charged residues and central hydrophobic region (Freire et al., 2013). The short hydrophobic C-terminal residues of C protein are cleaved by the viral serine protease (Amberg et al., 1994). The capsid protein forms dimers with charged surface on one side for binding to ssRNA genome and hydrophobic surface on the other for association with lipid membrane. The flavivirus

nucleocapsid lacks rigid well-ordered assembly structure. Both crystal structure and NMR structure of capsid protein are available (Dokland et al., 2004a; Ma et al., 2004). These structures showed that C-protein forms dimers which in turn arranged into tetramers with a highly positive charged arginine surface facing the inside of the tetramer core (Figure 1.11). These tetramers with a positive charged tunnel are speculated to be interacting with viral RNAs (Dokland et al., 2004b). In addition to its structural function, studies demonstrating C-protein translocation into the nucleus have been reported (Makino et al., 1989; Tadano et al., 1989; Westaway et al., 1997). It is possible that the C-protein acts as a histone mimic and interferes with cellular transcriptional machinery (Colpitts et al., 2011).

Figure 1.11. Flavivirus capsid protein structure. (A) C protein monomer. The Arginine rich surface and hydrophobic cleft are shown in boxes. (B) Dimer of C protein (C) Tetramers of C protein showing putative RNA binding tunnel in the middle and hydrophobic cleft facing outside. (PDB id: 1SFK) (Dokland et al., 2004a)

Membrane protein

The flaviviral membrane glycoprotein is about 165 amino acids long and contains two transmembrane helices (Mukhopadhyay et al., 2005). The hydrophobic C-terminal sequence of capsid protein guides the prM into ER lumen along with E protein (Mackenzie and Westaway, 2001). The M protein is translated as precursor M protein (prM) which is cleaved from capsid protein by ER signal peptidase (Murray et al., 1993). However, cleavage of prM

from C protein occurs only after the cleavage of C terminal of capsid protein to make sure the proper translocation of prM in ER lumen (Stocks and Lobigs, 1998) (Junjhon et al., 2008). The prM forms heterodimers with E protein on the ER lumens and protects the fusion loop of E protein from premature fusion during the transport of virion along the secretory pathway (Hsieh et al., 2014; Lorenz et al., 2002). The N-terminal precursor residues are cleaved by the host protease furin in the golgi to form mature membrane protein as part of viral maturation (Randolph et al., 1990). Along with the release of pr portion from prM, prM-E dimers dissociate and E-homodimers are formed (Yu et al., 2008).

Envelope protein

The envelope protein, the major antigenic component of flavivirus, is about 53 kDa and made up of transmembrane domain and ectodomains. The X-ray crystal structures of E protein ectodomain have been determined for TBEV, WNV, DENV2 and recently for ZIKV (Dai et al., 2016; Kanai et al., 2006; Modis et al., 2003; Nybakken et al., 2006; Rey et al., 1995b; Zhang et al., 2004). The ectodomains of these viruses show similar E protein homodimers with each monomer containing three domains. The N-terminus of the protein is in the central and called domain I. This domain I is flanked on one side by dimerization domain which contains fusion peptide while on the other by an immunoglobulin like domain containing receptor binding sites (Figure 1.12)

Figure 1.12. Structure of ectodomains from ZIKV E protein. (A) Schematic representation of E protein gene. The figure is adapted from (Dai et al., 2016) (B) Homodimer of sE protein. Individual domains for one monomer are labelled as domain I, domain II and domain III. The N- and C-terminus C-alphas are shown in red and blue spheres respectively. The fusion loop is shown in green box and labelled. PDB id: 5HJM (Dai et al., 2016)

Section 1.2.3: Non-structural proteins

Non-structural protein 1

Non-structural protein 1 is a 46 kDa protein comprising of three domains, the hydrophobic β roll, wing domain and central β ladder (Akey et al., 2014). The host signal peptidase cleaves NS1 from E protein while an unknown host ER protease cleaves the junction between NS1 and NS2A. The NS1 secondary structure is stabilized by six disulfide bonds formed by twelve conserved cysteines (Blitvich et al., 2001). NS1 undergoes post-translational N-linked glycosylation at two or three sites upon translocation to ER membrane (Jacobs et al., 2000; Winkler et al., 1989; Winkler et al., 1988). The cellular NS1 forms dimers and are membrane

associated (Winkler et al., 1989). The crystal structures of NS1 show NS1 in dimer form with two faces with different hydrophobicity. The “outerface” is hydrophilic with glycosylation sites exposed where the inner surface is hydrophobic (Figure 1.13) (Akey et al., 2014). The cellular NS1 has been shown to co-localize to RNA replication sites and can be trans-complemented although it does not have a catalytic mechanism in replication (Mackenzie et al., 1996). The mutations at the glycosylation site could be detrimental to RNA replication (Blitvich et al., 2001). NS1 interacts with NS4A during viral replication (Youn et al., 2012). NS1 is shown to be secreted from the cells via non-classical caveolin-dependent pathway and its extracellular concentration has been linked to onset of hemorrhagic fever in DENV. The secreted NS1 is in the lipid associated hexameric form and has been shown to interfere with the complement system (Chung et al., 2006).

Figure 1.13. Crystal structure of NS1 dimers from DENV. (A) Structure of NS1 dimer. A monomer is shown in grey and another in color. (B) A putative model of how NS1 could anchor to the membrane. The hydrophobic surface of NS1 formed by β roll domain would be facing the membrane. (C) Three domains of NS1 monomer. Figure adapted from (Akey et al., 2014) (PDB id: 4V0R)

Non-structural protein 2A and 2B

Non-structural protein 2A is a small hydrophobic membrane protein of about 26 kDa (Chambers et al., 1990a; Liu et al., 2006). As mentioned above, the NS1-2A cleavage is mediated by an unknown ER host protease in the ER lumen while the NS2A-NS2B cleavage is carried out by viral serine protease on the cytosolic side of the ER membrane indicating that NS2A could be transmembrane protein (Chambers et al., 1991; Chambers et al., 1990c; Falgout et al., 1991; Jan et al., 1995; Pethel et al., 1992). NS2A has been shown to interact with NS3 and NS3 and co-localize with the viral replication complex. DENV NS2A mutants are defective in viral RNA synthesis indicating a role for NS2A in viral replication (Xie et al., 2015). NS2A is involved in inhibition of IFN response as well with studies showing WNV NS2A mutants virus were attenuated *in vivo* (Liu et al., 2006). The viral protease can cleave NS2A internally at C-terminus yielding truncated form NS2A α of 20 kDa (Kummerer and Rice, 2002). The mutations at NS2A α block viral assembly in YFV while not affecting viral replication indicating that NS2A α is involved in assembly but not viral replication (Kummerer and Rice, 2002).

Non-structural protein 2B is also a small transmembrane protein of about 14 kDa (Falgout et al., 1991). The viral protease cleaves the NS2A-NS2B junction as well as the NS2B-NS3 junction. NS2B has N- and C-terminal hydrophobic regions for membrane anchorage and a central hydrophilic region which is the co-factor for NS3 protease. The co-factor region is 47 amino acids long and is necessary for NS3 protease folding and enzymatic activity. The

membrane associated residues of NS2B were mapped by NMR and HDX depicting the boundaries of NS2B-membrane association (Figure 1.14) (Li et al., 2015). The co-factor region alone is sufficient for NS3 protease activity *in vitro* although mutations in transmembrane regions of NS2B result in viral replication defects.

Figure 1.14. Membrane topology of DENV NS2B protein. The N- and C-termini are labelled. The border residues of the cofactor region are shown in red circles. The NS2B-NS3 cleavage recognition site residues are shown in magenta circles. Figure adapted from Li et al. (Li et al., 2015)

Non-structural protein 3

Non-structural protein 3 (NS3) is about 69 kDa and is essential in viral replication and polyprotein processing (Lescar et al., 2008). The N-terminal domain of DENV NS3 (residues 1-168) is a protease responsible for cleavage of polyprotein precursor into functional proteins while the C-terminal is an NTPase/helicase (residues 180-618) (Luo et al., 2015; Ma et al., 2008). The N-terminal protease is a serine protease with a catalytic triad comprising Asp, Ser and His at amino acid positions 75, 135 and 51 respectively (Chambers et al., 1990c). Early studies of NS3 revealed that NS3 protease alone was insoluble and required NS2B cofactor region for catalytic activity and proper folding (Yusof et al., 2000). The protein cleaves the

viral polyprotein at Capsid C, NS2A/2B, NS2B/3, NS3/4A and NS4B/5. The cleavage sites differ slightly between the virus species (Table 1.2). Mutagenesis studies of at the cleavage sites NS2A/2B, NS3/4A and NS4B/5 in YFV showed that different cleavage sites tolerate the mutations differently. For NS2A/2B cleavage site, the P1 mutations are not tolerated except for lysine while for NS3/4A and NS4B/5 sites, even the P1 Arg to Phe was tolerated. Instead, P1' site is more sensitive to mutations in NS3/4A site where changing from G to A results in inefficient cleavage. (Chambers et al., 1991; Chambers et al., 1993; Droll et al., 2000). Mutagenesis of conserved residues near WNV protease active site to DENV switched the substrate specificity of protease to DENV suggesting that protease and cleavage sites co-evolved. (Shiryaev et al., 2007). The *in vitro* studies have characterized the sequence of polyprotein cleavage depending on the cleavage efficiency of cleavage site mimic substrates/short peptides. It shows that in DENV2, NS3 protease cleaves in order of 4B/5 or 2A/2B, NS3/4A, NS2B/3 from most efficient to least (Chanprapaph et al., 2005; Niyomrattanakit et al., 2006). However, using FRET peptides, Gouvea et al., reported that DENV2 NS2B-NS3 protease cleavages the polyprotein sites in the order of C >> 2A/2B > 2B/3 > 4B/5 $\approx$ 3/4A from the most efficient to the least (Gouvea et al., 2007). For WNV, the protease cleavage is in order of 3/4A, 4B/5, 2A/2B followed by 2B/3 from most efficient cleavage to the least (Nall et al., 2004). It is possible that NS3 protease is regulating the cleavage efficiency depending on spatial closeness of the cleavage site to the protease since NS2B/NS3 cleavage has lowest catalysis efficiency in both WNV and DENV.

Table 1.2. Flavivirus polyprotein cleavage sites by NS2B-NS3 protease.

	Capsid C	NS2A/2B	NS2B/3	NS3/4A	NS4B/5
JEV	NKR/GGNE	NKKR/GWPA	TTKR/GGVF	AGKR/SAVS	SLKR/GRPG
WNV	KKR/GGTA	NRKR/GWPA	YTKR/GGVL	SGKR/SQIG	GLKR/GGAK
YFV	KRR/SHDV	FGRR/SIPV	GARR/SGDV	EGRR/GAAE	TGRR/GSAN
ZIKV	RKRR/GADT	SGKR/SWPP	TGKR/SGAL	AGKR/GAAL	VKRR/GGGT
DENV	RKR/STIT	ASRR/SWPL	KTQR/SGAL	SGRK/SITL	TPRR/GTGT

The crystal structures of NS3 protease covalently linked to NS2B cofactor region were solved for DENV, WNV and ZIKV (Aleshin et al., 2007; Erbel et al., 2006; Lei et al., 2016). These structures in free enzyme state and in complex with inhibitors show the chymotrypsin-like fold serine protease with two β barrels (Erbel et al., 2006; Hammamy et al., 2013; Lei et al., 2016; Noble et al., 2012; Robin et al., 2009). The NS2B cofactor N-terminus contributes to first β barrel forming β sheet, while the C-terminus is flexible in free enzyme structures (Figure 1.15B). In substrate bound structures, the NS2B C-terminus folds back into a β hairpin loop contributing the substrate binding site (Figure 1.15C). Unlike HCV NS3 protease, the NS2B cofactor plays a role in substrate recognition in addition to protein stabilization (Figure 1.15A-C).

Figure 1.15. Structures of NS3 protease from HCV and WNV protease. (A) HCV NS3/4A protease (PDB id: 1A1R) (Kim et al., 1997) (B) WNV NS2B₄₇NS3 protease in free enzyme state (PDB id: 2GGV) (Aleshin et al., 2007) (C) in complex with a peptidic inhibitor (PDB id: 2FP7) (Erbel et al., 2006). The NS2B or NS4A cofactor is color in magenta. HCV NS3 protease is colored in cyan and WNV NS3 protease in yellow. The catalytic triads are shown in sticks. The Zn²⁺ ion is shown as a grey sphere in HCV NS3/4A protease.

The crystal structures of the helicase domain have been reported for JEV, YFV, MVEV, DENV, Kunjin, Kokobera virus and recently ZIKV (Mastrangelo et al., 2007; Tian et al., 2016b; Wu et al., 2005; Xu et al., 2005). The structures show the C-terminal helicase contains 3 domains (Figure 1.16) with similar overall fold with RMSD ranging from 0.75-1.9 Å. Domain I and II adopt the RecA like fold with 8 conserved motifs and is essential for RNA binding and adenosine triphosphate (ATP) hydrolysis (Figure 1.16A). The ATP/ nucleotide triphosphate (NTP) binding site is located between the domain I and II (Xu et al., 2006). The conserved Walker A, also known as P-loop and Walker B motifs are responsible for

recognizing the NTPs and binding of cations (Mg^{2+} or Mn^{2+}) (Figure 1.16A)(Sampath et al., 2006; Wu et al., 2005). Domain III sits on top of the ssRNA binding tunnel and is involved in RNA binding (Sampath et al., 2006; Wu et al., 2005; Xu et al., 2006). The domain II interacts with domain III via a long β hairpin (Wu et al., 2005). The helicase domain has both RNA-stimulated NTPase activity and ATP-dependent RNA unwinding activity (Benarroch et al., 2004). Currently only DENV and ZIKV helicases structures in complex with ssRNA have been reported (Chen et al., 2016; Luo et al., 2008b). In both structures, the NS3 helicase translocate from 3' to 5' direction with subdomain I binding to the 3' end of ssRNA and with subdomain II interacting with 5' end. In DENV-ssRNA structure, RNA binding is mediated via interactions between the sugar phosphate backbones whereas for ZIKV-ssRNA structure, more interactions are observed between the individual bases and helicase residues compared to that of DENV structure (Luo et al., 2010; Tian et al., 2016b). These structures also showed different conformational changes upon ssRNA binding with domain II and III rotating in different axis (Figure 1.16C). Both DENV and ZIKV helicases have been captured in complex with ATP, adenosine diphosphate (ADP) with inorganic phosphate (Pi) and with manganese ion (Mn^{2+}) and adenylyl imidodiphosphate (AMP-PNP) – Mn^{2+} as well (Cao et al., 2016; Jain et al., 2016; Tian et al., 2016a; Wu et al., 2005; Xu et al., 2005). In DENV – ADP- Pi - structure, the P loop at the NTP binding site moved inwards making the ATP binding site smaller/more restricted compared to the apo structure. It has been reported that NS3 helicase interacts with NS5 and NS4A although it is not clear how different enzymatic activities of these viral proteins are coordinated (Chen et al., 1997a; Kapoor et al., 1995; Tay et al., 2015).

Figure 1.16. Crystal structures of NS3 helicase domains from flavivirus (A) Crystal structure of NS3 helicase domain from JEV (PDB id: 2Z83) showing domain I, II and III. (B) Crystal structure of NS3 helicase domain from DENV4 in apo form (Grey) and in complex with ssRNA₁₂. The helicase-ssRNA structure is shown with color domains, domain I (magenta), II in blue and III in green. (PDB id: 2JLQ and 2JLW respectively) (C) Crystal structure of NS3 helicase domain from ZIKV in apo form (grey) and in complex with ssRNA (colored domains) (PDB id: 5GJB and 5GJC respectively). (D) Proposed mechanisms of domain movements upon ssRNA binding to DENV and ZIKV helicases. Panel D is adapted from (Tian et al., 2016b)

The full length NS3 structures showed distinct domain arrangements between protease and helicase compared to that of HCV (Figure 1.17) (Assenberg et al., 2009; Luo et al., 2008a; Yao et al., 1999). While the HCV NS3 full length structure shows a globular conformation with protease domain to the side of helicase domain, the flavivirus NS3 structures shows an elongated conformation with helicase on top of protease (Figure 1.17). The crystal structures of DENV4 full length NS3 covalently linked to NS2B cofactor N-terminus 18 residues (DENV4 NS2B₁₈NS3) were solved by Luo et al., (Luo et al., 2008a) (Luo et al., 2010) (Figure 1.17B,C). Although 18 amino acids from NS2B cofactor region render NS3 to be soluble, the protein is not catalytically active (Erbel et al., 2006; Luo et al., 2008a). The full length structure in complex with ADP-Mn²⁺ shows a slightly different conformation where the protease domain is rotated and no longer blocking the ATP binding site (Figure 1.17D). Although the linker between the protease and helicase domain is somewhat flexible, Luo et al., showed that mutagenesis of a proline residue to glycine in the linker region between helicase and protease domain lowered the helicase activity in both *in vitro* and cell culture assays (Luo et al., 2010). This demonstrated that introducing more flexibility to the linker might allow higher numbers of possible conformations which are not optimal for viral replication. (Mastrangelo et al., 2007). The full length NS3 with NS2B cofactor 47 residues from MVEV has been reported by Assenberg et al., (Figure 1.17B) (Assenberg et al., 2009). MVEV NS2B₄₇-NS3 too adopts the extended conformation similar to the DENV4 full length structure. However, full length NS2B₄₇-NS3 from MVEV showed distinct conformation and domain arrangement from NS2B₁₈-NS3 (Luo et al., 2008a) where protease domain is shifted away from ATP binding pocket of helicase domain (Figure 1.17D). No crystal structure of full length NS3 in complex with substrates/inhibitors in protease domain has been reported. The enzymatic activities of full length NS3 have been compared to that of each domain alone and the results have been somewhat controversial. The WNV protease domain alone has 2

fold higher k_{cat} and similar K_m compared to the WNV NS3 full length (Chappell et al., 2007) although in another study, WNV protease has similar k_{cat} with two fold higher K_m compared to full length construct (Chernov et al., 2008). Chernov et al., also reported that WNV NS3 full length has better ATPase activity compared to that of WNV helicase construct but full length protein failed to unwind dsDNA while the helicase domain can (Chernov et al., 2008). In full length DENV4 NS3, both helicase and ATPase activities are enhanced compared to the helicase domain alone reported by Luo et al., (Luo et al., 2008a). However, Xu et al., reported that NS3 helicase has higher ATPase activity compared to full length (Xu et al., 2005). Mastrangelo et al., reported the similar ATPase/RTPase activities of Kunji NS3 helicase and full length but increased RNA unwinding activity in full length NS3 (Mastrangelo et al., 2007). Assenberg et al., reported similar ATPase and helicase activities for the MVEV helicase and full length NS3 (Assenberg et al., 2009). It is possible that the discrepancy between the enzymatic activities of the helicase and of the full length could be specie specific factors, different construct lengths and/or different assay protocols (eg. ATPase activity was either measured using malachite green or NADH coupled ATPase assay). The role of NS3 in antagonizing innate immunity has been established and mentioned in the previous section. The NS3 roles in replication that are unrelated to its enzymatic reactions have been demonstrated via mutagenesis and reverse genetics studies. The NS3 hydrophobic loop has been shown to be involved in viral replication but not in polyprotein processing by Matsuan et al., (Matusan et al., 2001). Being an essential protein for viral replication, assembly, and viral immune evasion makes NS3 an attractive drug target. More details on structural and functional properties of NS3 protease are discussed in Chapter 1.4.

Figure 1.17. Full length crystal structures of HCV (hepacivirus), MVEV and DENV (flavivirus). The NS4A/NS2B cofactor is shown in magenta. The protease domain is shown in a square box. (A) HCV NS3 full length structures shows globular fold (PDB id: 1CU1) (Yao et al., 1999). (B) MVEV NS3 full length with 47 residues of NS2B cofactor. The protease domain is rotated away from the helicase domain. (PDB id: 2WV9) (C) DENV4 NS3 full length structure with 18 residues of NS2B cofactor N-terminus. The protease domain sits underneath the ATPase binding site. (PDB id: 2VBC) (D) DENV4 NS2B₁₈NS3 structure in complex with ADP-Mn²⁺ (PDB id: 2WHX). In this structure, the protease domain rotated and is no longer right beneath the ATP binding site. Both MVEV and DENV4 NS3 full length structures are in elongated shape in contrast to HCV.

NS4A and NS4B proteins

Similar to NS2A and 2B, NS4A and 4B are small hydrophobic proteins of about 16 kDa and 27 kDa (Chambers et al., 1990a; Stern et al., 2013). The NS4A and 4B co-localise with replication complexes and have been shown to be involved in ER membrane arrangement during replication (Zou et al., 2014). The junction between NS4A and 4B is cleaved by NS3 protease (Cahour et al., 1992). NS4A has been shown to repress the helicase activity *in vitro* (Shiryaev et al., 2009). The role of NS4A and 4B in inhibition the type I IFN signalling has

been studied. Interactions between NS1 and NS4A have also been established (Munoz-Jordan et al., 2005; Munoz-Jordan et al., 2003).

NS5 protein

NS5 is the largest protein in the genome and most conserved as well (Lescar, 2012). The NS5 is about 103 kDa and contains at least two distinct enzymatic activities, MTase activity and RdRp activity (Bollati et al., 2010; Egloff et al., 2007; Issur et al., 2009; Yap et al., 2010; Zhao et al., 2015a). The N-terminal domain is a SAM dependent MTase responsible for 5' capping of genome (Egloff et al., 2007; Issur et al., 2009; Yap et al., 2010). Crystal structures of MTase domain were determined for DENV, WNV, YFV and ZIKV (Coloma et al., 2016; Geiss et al., 2009; Jain et al., 2017; Zhang et al., 2017). The MTase domain is globular with three sub domains (Coloma et al., 2016; Geiss et al., 2009; Jain et al., 2017; Zhang et al., 2017). Subdomain 1 consists of the GTP-binding pocket, while subdomain 2 is responsible for S-Adenosyl methionine (SAM) binding and MTase activity (Figure 1.18B). The RNA binding site is located at the positively charged groove formed between the GTP-binding pocket and the SAM-binding pocket (Zhao et al., 2015b). Subdomain 3 is the smallest domain and proposed to coordinate the methylation between first guanine N7 methylation and 2'O methylation (Figure 1.18B). The C terminal NS5 is the RdRp domain (Chambers et al., 1990a). The flaviviral RdRp domain has similar cupped right hand architecture with globular fold containing fingers, palm and thumb domains. Two zinc binding sites are located in the thumb and fingers subdomains. The fingers and thumb domains form the RNA binding tunnel (Figure 1.18) and are responsible for switching of NS5 activity from *de novo* initiation to elongation (Tian et al., 2016b; Zhao et al., 2015a; Zhao et al., 2015b). The palm domain is the most conserved domain of RdRp with two anti-parallel β strands surrounded by eight α helices and responsible for binding of metal ions, nucleotides and RNA as well as catalytic phosphoryl transfer. Apart from these essential catalytic activities for RNA replication, NS5

is also involved in immune evasion (Miorin et al., 2017). The NS5 contains a nuclear localization signal (NLS) and localizes in the nucleus during the DENV infection in a strain specific manner (Tay et al., 2016). TBEV NS5 inhibits IFN signaling by binding to JAK1 and Tyk2 and inhibiting their phosphorylation, inactivating the downstream activation via STAT1 pathway (Miorin et al., 2017; Miorin et al., 2013).

Figure 1.18. (A) Full length NS5 structure of DENV3 showing MTase domain on top and RdRp below. (B) MTase domain of DENV with Cap0-RNA. The GTP binding site, SAM binding sites and cap0-RNA binding sites are indicated. (C) DENV NS5 RdRp domain showing the RNA binding tunnel and priming loop. (Figures are adapted from (Zhao et al., 2015a; Zhao et al., 2015b) (PDB id: 4V0R and 5DTO)

Chapter 1.3: Immune response and evasion strategies by flavivirus

Innate immunity

The innate immunity detects the viral infections by recognizing the pathogen-associated molecular patterns (PAMPs) via pattern recognition receptors (PRRs) in the cell. Major PRRs involved in detecting the viral infections are (1) retinoic inducible gene-I (RIG-I) like receptors (RLRs), (2) Toll-like receptors (TLRs), (3) nucleotide binding oligomerization domain (NOD)-like receptors (NLRs), (4) cytosolic DNA sensors, (5) C-type lectin receptors (Kanneganti, 2010; Kell and Gale, 2015; Nasirudeen et al., 2011) (Figure 1.19).

Figure 1.19: Major PAMPs recognition receptors and the downstream pathways. Figure adapted from (Shekarian et al., 2017).

RLRs consist of RIG-I, Melanoma differentiation associated protein 5 (MDA5) and Laboratory of Genetics and Physiology 2 (LGP-2). RIG-I recognizes short double stranded RNAs (dsRNAs) with 5' triphosphate while MDA5 recognizes longer dsRNAs (Feng et al., 2012; Kell and Gale, 2015). The activation of RLRs leads to a type I interferon response with the secretion of IFN- α/β where they activate the anti-viral interferon stimulated genes (ISGs) in neighboring cells (Nasirudeen et al., 2011). Incubation of cells with type I interferons prior to viral infection inhibits viral replication *in vitro*, whereas post infection treatment is not effective. (Best et al., 2005b; Diamond et al., 2000; Lin et al., 2004b; Samuel and Diamond, 2005). Since activation and secretion of IFN are crucial for viral infectivity, flaviviruses must come up with strategies to antagonize the RIG-I/MDA5 pathway. Various studies have been

done to discover the viral proteins and targeted immune proteins that are involved in immune evasion of flavivirus (Munoz-Jordan et al., 2003). Flaviviral genome is capped at 5' end by methyltransferase domain of NS5 (Lindenbach and Rice, 2003b). WNV virus lacking 2'-O-Methyltransferase activity is sensitive to type I IFN and replicate poorly in both *in vitro* cell culture and in mice (Daffis et al., 2010; Szretter et al., 2012). DENV NS4A has been shown to associate with MAVS making it unavailable for RIG-I binding and thus inhibiting the IRF3 activation. Both DENV and WNV NS2B3 bind to 14-3-3 ϵ , a protein responsible for RNA-bound RIG-I translocation to mitochondria therefore preventing RIG-I-MAVS interactions for downstream signaling (Chan and Gack, 2016; He et al., 2016; Liu et al., 2012). In addition to active interference of viral proteins in inhibiting Type I IFN production, flavivirus escape the detection by RLRs by replicating in convoluted vesicle packets in ER membrane shielding the newly synthesized RNA from detection (Espada-Murao and Morita, 2011; Overby et al., 2010). This may result in the observed delay of the IFN- α/β response during the flaviviral infections (Espada-Murao and Morita, 2011; Overby et al., 2010). The toll-like receptors (TLRs) are other important PRRs against viral infections (Alexopoulou et al., 2001). TLRs found in the endosomal vesicles can recognize both dsRNAs and ssRNAs and induce a type I IFN response and cytokine production (Adachi et al., 1998; Tsai et al., 2009). WNV NS1 might inhibit TLR-3 in cell type dependent manner (Baronti et al., 2010; Wilson et al., 2008) whereas WNV E protein inhibits downstream NF- κ B signaling triggered by activated TLRs (Arjona et al., 2007). During the viral genome synthesis, flaviviruses produce non-coding, subgenomic RNAs of about 300-500 bases from the 3' UTR of viral genome (Lin et al., 2004a). These RNAs are called subgenomic flaviviral RNAs (sfRNAs). Schuessler et al., showed mutant WNV deficient in sfRNA production replicated poorly in mice indicating the role of these sfRNAs in immune evasion by the virus (Schuessler et al., 2012). Manokaran et al., also showed that the sfRNAs from DENV binds to TRIM25

(tripartite motif protein 25), an important regulator of PRR signaling and inhibit the RIG-I activated downstream signaling (Manokaran et al., 2015). Activation of RIG-I/MDA5 or TLRs trigger MAVS or NF- κ B pathway leading to phosphorylation of Interferon response factor3 (IRF-3) and IRF-7. The phosphorylated IRF-3 and IRF-7 enter the nucleus and activate the expression of type I interferons. Two kinases named IKK ϵ and TBK-1 responsible for IRF-3 phosphorylation are also inhibited by NS2B3, NS2A, NS4A and NS4B of flaviviruses antagonizing the IFN-I production (Anglero-Rodriguez et al., 2014; Dalrymple et al., 2015; Liu et al., 2004; Liu et al., 2006). In addition to RLRs and TLRs, the cytosolic PAMPs are recognized by Nod like receptors (NLRs) (Inohara et al., 2005). The NLRs are responsible for a wide range of cytosolic PAMPs from bacteria, viruses and fungi. Upon binding of PAMPs, the NLRs form inflammasome. The inflammasome formation activates caspase-1 and subsequently results in secretion of proinflammatory cytokines (Kanneganti, 2010). Role of NLRs in the flaviviral infections were only recently discovered and viral components that are responsible for activation of NLRs are not yet known. During WNV infection, mice deficient in NLRP3 or caspase-1 showed similar phenotype as mice lacking interleukin-1 receptor indicating that NLRP3-inflammasome pathway was important in IL-1 β production (Ramos et al., 2012). The caspase-1 inflammasome pathway was found to be dependent on Apoptosis-Associated Speck-Like Protein Containing A CARD (ASC) during WNV infections (Kumar et al., 2013). Although a number of viruses such as influenza A virus (IAV) target IL-1 β production, the evasion strategies of flaviviruses on NLR pathway have not been established yet. The cytosolic DNA sensors are also part of PRRs and detect foreign DNA. One of the important cytoplasmic DNA sensors is the stimulator of IFN genes (STING) (Ishikawa and Barber, 2008; Ishikawa et al., 2009). Activation of STING leads to activation of IRF-3 and NF- κ B pathway. DENV NS2B-NS3 protease cleaves STING inactivating the IFN response (Aguirre et al., 2012).

During viral infection, following activation of type I IFN response, the secreted IFN- α/β binds to interferon response alpha receptor 1 and 2 (IFNAR1 and IFNAR2), which activates the JAK/STAT pathway leading to the expression of ISGs (Figure 1.20B) (Elong Ngonu and Shresta, 2018; Ivashkiv and Donlin, 2014). Flaviviruses employ various strategies to mitigate the activation of the JAK/STAT pathway. The flaviviral protein NS5 has been shown to antagonize the JAK/STAT pathway in DENV, WNV, ZIKV, TBEV and YFV infections by binding to STAT2 and interfering with STAT2/TYK2 phosphorylation (Ashour et al., 2009; Best et al., 2005a; Grant et al., 2016; Jones et al., 2005; Kumar et al., 2016; Mazzon et al., 2009; Morrison et al., 2013; Werme et al., 2008). NS2A, NS4A, and NS4B also interfere with the JAK/STAT pathway by reducing phosphorylation and nuclear translocation of STAT1 (Guo et al., 2005; Lin et al., 2004b; Liu et al., 2005; Munoz-Jordan et al., 2005; Munoz-Jordan et al., 2003).

The other two important branches of innate immunity are the complement system and natural killer cells (NK cells). The complement system bridges the innate immunity with adaptive immunity. The NS1 protein from various flaviviruses secreted into the extracellular matrix has been shown to interfere with the complement system (Avirutnan et al., 2011; Avirutnan et al., 2006; Kurosu et al., 2007). NK cells lyse infected cells either by releasing cytotoxic granules or by inducing apoptosis. Flaviviruses down regulate the lytic activity of NK cells by upregulating the major histocompatibility complex I (MHC-I) (Cheng et al., 2004; Hershkovitz et al., 2008; Lobigs et al., 2004). The flavivirus infection seems to delay the activation of apoptosis as well although the viral proteins that are involved in the process are not identified yet (Lee et al., 2005; Medigeschi et al., 2007). Figure 1.20 shows an overview of the strategies employed by flaviviruses to antagonize the innate immune response.

A

B

Figure 1.20. IFN antagonism strategies by Flaviviruses. (A) The strategies applied by different flaviviruses to inhibit the IFN production. (B) The IFN are released into cytosol where they bind to IFNARs activating the downstream antiviral pathway involving ISGs. Various flavivirus non-structural proteins are involved in antagonizing the JAK/STAT pathway. Figure adapted from (Miorin et al., 2017)

Adaptive immunity

Adaptive immunity is made up of humoral immunity involving antibodies and cellular immunity involving CD4⁺ and CD8⁺ cells (Elong Ngonu and Shrestha, 2018). The major components of humoral immunity are neutralizing antibodies (Abs) which recognize pathogen specific molecular patterns. In the case of flaviviral infections, the recognized epitopes are mostly located on the viral E-glycoprotein and to a lesser extent on NS1 protein (Muller and Young, 2013). The major determinant of viral evasion from adaptive immunity relies on low-fidelity RNA-dependent-RNA-polymerase which generates quasispecies of viruses which can evade recognition by antibodies (Elong Ngonu and Shrestha, 2018). DENV has been found to impair the upregulation of MHCs and co-stimulatory molecules in monocytes (Halstead, 2014). Similarly, JEV infections interfere with TLR pathways leading to poor CD8⁺ T-cell responses (Jiang et al., 2014; Nazmi et al., 2014). During the maturation of virus, cleavage of prM to E protein is necessary. This process is not efficient leading to a mixed population of mature and immature viral particles (Junjhon et al., 2010). Although the detailed mechanisms are not clear, the presence of immature viral particles is employed as one of the strategies to escape from antibody neutralization (Rey et al., 2017; Rodenhuis-Zybert et al., 2010).

Chapter 1.4: Diagnostic methods, antivirals and vaccines for flavivirus infections

Section 1.4.1: Diagnostic methods

The current available diagnostic methods for flavivirus infections are either based on isolated virus from patient samples or isolated serum/antibodies from the patient samples (Shu and Huang, 2004). Isolation of DENV is achieved by inoculation of specimen in cell culture or into mosquitoes or by injection into mice. Upon isolation of the virus from the culture, the presence of DENV is confirmed by specific antibodies. The virus isolation method is reliable and provides the possibility to propagate the virus for downstream studies. However, the whole process of virus isolation is time-consuming and expensive (Peeling et al., 2010; Shu and Huang, 2004). Flavivirus can be detected by extraction of viral genome from the patient samples and subsequent detection using PCR based methods such as real-time PCR (RT-PCR), quantitative PCR (qPCR) or nested PCR assays or antibody based methods such as ELISA or immunochromatographic assays (Blacksell, 2012; Hunsperger et al., 2014; Pal et al., 2015). These methods are sensitive and cost-effective and can also be applied to detect virus specific antibodies found in the patient serum during the infection. In fact commercially available kits for DENV detection are based on immunochromatographic detection of both NS1 and IgG/IgM (Hunsperger et al., 2014). A number of PCR assays detecting across the pan-flaviviruses have been reported in the recent years (Blacksell, 2012; Peeling et al., 2010). Other serological tests for flaviviral infections include haemagglutination-inhibition tests, neutralization tests and dot-blot assays (Jusoh and Shueb, 2017). Again, the sensitivity of serological based diagnostic tools are complicated by the cross reactivity of antigens among flaviviruses (Muller and Young, 2013; Shan et al., 2015). The combination of both viral genome and viral specific antibodies detection is currently the best strategy to maximize the sensitivity of diagnostic tests (Jusoh and Shueb, 2017). The new detection methods using

microfluidics, piezoelectric, flow cells sensors and microarrays are in development (Blacksell, 2012).

Section 1.4.2: Vaccines against flavivirus

Vaccines are available for YFV, TBEV and JEV. Currently available YFV vaccines are based on the live attenuated virus (LAV)s (Ishikawa et al., 2014). There are six YFV vaccines in the market approved by WHO and all of them are based on LAVs 17D strain (Ishikawa et al., 2014). For JEV, there are both LAVs based vaccines and inactivated virus vaccine. For TBEV vaccines, the vaccines are available depending on the region and prevalent TBEV sub type (Ishikawa et al., 2014). There are currently no human vaccines against WNV, DENV and ZIKV.

The first DENV vaccine, Dengvaxia© by Sanofi Pasteur, was approved in 2016. It is a live-attenuated chimeric vaccine where YFV-17D strain was fused with prM and E proteins from DENV 1,2,3 and 4 viruses (Capeding et al., 2014; Guy et al., 2011; Villar et al., 2015). The vaccine is administered in three doses at six months intervals (Capeding et al., 2014). The efficacy of Dengvaxia© is varied among the four serotypes with the efficacies for DENV 1 and 2 dipping to 35-50% (Villar et al., 2015). It also has been reported that naïve individuals have higher chance of developing severe dengue after the vaccination (Aguiar et al., 2017; Aguiar et al., 2016). Currently there are two more vaccines in the Phase III clinical trials (Kirkpatrick et al., 2015; Sirivichayakul et al., 2016). Other vaccines in the pipeline using different vaccine development strategies, such as inactivated virus, live attenuated virus (LAV) or recombinant subunit, or DNA vaccine are listed in Table 1.3 (Morabito and Graham, 2017). The LAV is obtained by serial passaging the wild type virus in cell culture, resulting in replication impaired virus with sufficient immunogenicity. This vaccine strategy is risky due the possibility of the attenuated virus reverting back to the virulent form. The LAV vaccines are usually administered in a series of doses in 3 – 6 months intervals. Another

LAV which entered Phase III clinical trials called DENVax by Takeda has shown seroconversion in all four DENVs after two doses (Sirivichayakul et al., 2016). This vaccine is based on attenuated virus PDK-53 with chimeric strains of prM and E genes from DENV 1,3 and 4 (Sirivichayakul et al., 2016). TV003, another LAV candidate in Phase III trials, is attenuated by deletion of 3'UTR region and contained prM and E genes from DENV 1,3 and 4 (Durbin et al., 2013a; Kirkpatrick et al., 2015). Other vaccines that are in phase I trials include DEN-80E by Merck, DIME-VR-P by Naval Medical Research Centre and TDENV-PIV by Walter Reed Army Institute of Research and Glaxo Smith Kline (Low et al., 2017). DEN-80E and DIME-VR-P are recombinant subunit vaccines (Manoff et al., 2015). The recombinant subunit vaccines are made by recombination of fragments or full length DENV structural proteins into a plasmid. These vaccines show poor immunogenicity and require adjuvant to elicit sufficient immune response (Holbrook, 2017b; Low et al., 2017; Morabito and Graham, 2017). TDENV-PIV is an inactivated virus vaccine which is produced from viral particles that are inactivated using chemicals such as formalin and are thought to be safer than LAVs (Beckett et al., 2011).

Since WHO declared ZIKV as Public Health Emergency of International Concern in 2015, there are currently more than 40 vaccine candidates against ZIKV in preclinical development, five in Phase I trials and two in Phase II trials (Lin et al., 2018; Morabito and Graham, 2017; Weaver et al., 2016). Similar to DENV vaccines, ZIKV vaccines target prM and E proteins and can be grouped into live attenuated virus vaccines, inactivated virus vaccines and nucleic acid vaccine. The two vaccines in Phase II trials, VRC5283 and mRNA-1325 are both nucleic acid vaccines (Dowd et al., 2016; Richner et al., 2017). VRC5283 by the Vaccine Research Centre, National Institute of Allergy and Infectious Diseases (NIAID) is a DNA based vaccine which encodes the full ZIKV prM-E sequence (Dowd et al., 2016). The mRNA-1325, developed by Moderna Therapeutics, is an mRNA vaccine with mRNA from

Micronesia 2007 ZIKV strain encapsulated in a lipid nanoparticle (Richner et al., 2017). Other vaccine candidates in Phase I trials are listed in Table 1.3. It is to be noted that cross reactivity and antibody enhancement of ZIKV infections by previous flavivirus infections have been reported recently and should be considered as a factor during vaccine development (Lin et al., 2018; Morabito and Graham, 2017).

Table 1.3. Flavivirus vaccines currently under development				
Vaccine	Type	Developer	Current stage	References
WNV				
rWN/DEN4 Δ30	Live attenuated	NIAID	Phase I	(Durbin et al., 2013b; Pierce et al., 2017; Pletnev et al., 2006)
ChimeriVax-WN02	Live attenuated	Sanofi Pasteur	Phase II	(Arroyo et al., 2004; Monath et al., 2006)
HydroVax-001	Inactivated	Oregon Health & Science	Phase I	(Pinto et al., 2013)
HBV-002	Recombinant subunit	Hawaii Biotech	Phase I	(Lieberman et al., 2009)
VRC 302/VRC 303	Nucleic acid	Vical/NIAID	Phase I	(Ledgerwood et al., 2011)
DENV				
Dengvaxia (CYD-TDV)	Live attenuated	Sanofi Pasteur	Licensed	(Capeding et al., 2014; Guy et al., 2011; Villar et al., 2015 {Guy, 2011 #856})
TV003	Live attenuated	NIAID	Phase III	(Durbin et al., 2013a; Kirkpatrick et al., 2015)
DENVax	Live attenuated	Takeda	Phase III	(Sirivichayakul et al., 2016)
TDEV-PIV	Inactivated	GSK, Firocruz & WRAIR	Phase I	(Schmidt et al., 2017)
V180	Recombinant subunit	Merck	Phase I	(Coller et al., 2011)
TVDV	Nucleic acid	U.S. Naval Medical Research Centre	Phase I	(Beckett et al., 2011)
ZIKV				
ZPIV	Inactivated virus	WRAIR/NIAID	Phase I	(Abbink et al., 2016; Larocca et al., 2016)
MV-Zika	Vector based	Themis Bioscience	Phase I	(Ramsauer et al., 2017)
VRC5283	Nucleic acid	NIAID	Phase II	(Dowd et al., 2016)
VRC5288	Nucleic acid	VRC/NIAID	Phase I	(Dowd et al., 2016)
GLS-5700	Nucleic acid	GLS-5700	Phase I	(Muthumani et al., 2016)
mRNA-13	Nucleic acid	Moderna Therapeutics	Phase I	(Richner et al., 2017)

Antivirals against flavivirus

The antivirals targeting flaviviral infections can be separated into direct acting antivirals, which target the viral proteins and host-targeting antivirals which target the cellular factors involve in viral propagation. In the case of viral targets, the target viral protein should be essential for viral replication and have low rate of allowed mutations. Three most common targeted viral proteins are (1) non-structural protein 5, (2) non-structural protein 3 and (3) envelope protein. The first two proteins carry out the major enzymatic activities required for the viral replication and polyprotein processing, while E-protein is involved in attachment, entry and maturation. Owing to its low fidelity RdRp, flavivirus have high mutation rates. This results in intragenetically varied quasispecies during the infection which can give rise to drug resistance. Combination of antivirals against more than one viral or host target should be applied to prevent the rise of drug resistance. In this section, we will review briefly regarding the development of antivirals against non-structural protein 5 and non-structural protein 3. The inhibitors against non-structural protein 3 are discussed more in details in Chapter 1.5.

The NS5 protein is the most conserved protein in flaviviral genome and hence, an attractive drug target. The NS5 is a multifunctional protein with N-terminal methyltransferase and C-terminal RdRp. On top of methyltransferase activity, the NS5 N-terminal domain has guanylyltransferase activity with a GTP binding site (Issur et al., 2009). High throughput assays have identified two nucleosides analogues that bind to the GTP binding site and inhibit WNV replication in cell culture (Chen et al., 2013). During the N7 and 2'O methylation, the N-terminal MTase uses SAM as methyl group donor (Egloff et al., 2007). The viral screening campaign of compounds against the methyltransferase activity has identified two compounds that inhibited the methyltransferase activity of WNV, DENV and YFV in *in vitro* biochemical assays and reduced viral titres for WNV, DENV and JEV (Brecher et al., 2015). Structural studies of DENV-3 and WNV MTase revealed a conserved hydrophobic cavity near SAM binding site and a small molecule inhibitor bound to DENV MTase was identified

(Dong et al., 2010; Lim et al., 2011). The C-terminus of NS5 is a RdRp. The NS5 RdRp is the most conserved part of the genome and is distinct from the eukaryotic polymerases making it one of the most sought after drug-targets. Major RdRp inhibitor compounds are nucleoside inhibitors, substrate analogues. These compounds enter the cells and subsequently are phosphorylated by cellular kinases. After phosphorylation, the compounds inhibit RdRp activity via chain termination or lethal accumulation of mutations. A few 2'C-methylated nucleosides have been reported to be broad spectrum inhibitors against flaviviruses *in vitro* biochemistry and cell culture assays. However, none of the leads were pursued further than Phase I trials due to low efficacy or cellular toxicity. Requirement to be phosphorylated by cellular kinases complicates the efficacy of the compounds in *in vivo* studies due to substrate specificity of different cellular kinases depending on tissue tropism while phosphorylated prodrugs have poor permeability and are unable to cross the cellular membranes. Another issue of nucleoside inhibitors is the binding to off-target polymerases resulting in toxicity especially to the mitochondrial DNA polymerase- γ . Allosteric inhibitors of DENV RdRp were reported in biochemical assays and in *in vitro* cell based assays. These inhibitors target the N-pocket near the active site and prevent the transition of RdRp from initiation to elongation state. One issue that could arise from the allosteric inhibitors is the low genetic barrier to resistance as is the case for allosteric inhibitors of HCV polymerase.

The NS3 is another multifunctional protein with N-terminal protease and C-terminal helicase. With the success of protease inhibitor drugs for HIV and HCV, the NS3 protease domain of flaviviruses has also been the drug target of many high throughput screening campaigns. Similar to HCV protease, the flaviviral protease has a shallow, solvent exposed active site which makes the discovery of inhibitors challenging. The flaviviral protease recognizes strong dibasic or polybasic residues. The interactions between the substrate and the protease residues are relatively conserved across the different species, and therefore could be a target

for broad-spectrum antivirals against flaviviral infections. Due to its preference for polar substrates, there is often a large difference between the inhibition efficiency of biochemical assays and cell based assays due to the poor permeability of the polar substrates/inhibitors. Many NS3 inhibitors are substrate mimicking compounds that binds to the active site of NS3. Studies have tried to improve the cell permeability and inhibition efficiency by modifying the N- and C-terminus of peptidomimetic inhibitors and incorporating non-natural amino acids at P1 and P2 residues. Allosteric inhibitors that disrupt the interaction between the NS2B cofactor and NS3 have also been reported. The current developments against the NS3 protease inhibitors are discussed more in detail in chapter 1.5. The C terminal of NS3 is a helicase/NTPase domain with RNA binding tunnel/groove and NTP binding site. Due to its similarity to the cellular helicases, these pockets impose the problem of helicase/NTPase inhibitors binding to the cellular proteins. A benzoxazole analogue with low cytotoxicity was reported to inhibit replication of DENV and YFV in cell cultures and reduce viral load in DENV infections in mice (Byrd et al., 2013). Ivermectin, an approved drug to treat parasitic infections, was shown to be inhibitory on flaviviral helicases. Screening of fragment based inhibitors of NS3 helicase was not successful (Coutard et al., 2014).

Chapter 1.5: X-ray crystallography in drug-discovery

The field of X-ray crystallography has grown much during the past 15 years. The Protein Data Bank (PDB) now holds more than 100,000 macromolecular structures with 80% of them determined in the past decade (Shi, 2014). A typical experimental flow of 3D structure determination of a protein is shown in Figure 1.21 (Abola et al., 2000). The most common bottlenecks in crystallography are obtaining the soluble homogenous protein of interest, finding the crystallization condition, and obtaining the crystals that have the strong diffraction pattern that can be detected. In general, the solubility of protein could be improved by designing the recombinant protein with solubility tags such as SUMO while the protein purification is enhanced using high pressure liquid chromatography equipment such as AKTA (Joachimik, 2009; Tickle et al., 2004). Automated approaches in initial crystallization screening by liquid handling robots such as PHENIX / Mosquitoes combined with 96 wells screening plates have dramatically reduced the amount of protein required for screening trials (Newman et al., 2005). With the automated approaches in crystallography, efficient crystallization screens have been developed by a few companies such as Hamptons Research, Molecular Dimensions, and Jena biosciences (Joachimik, 2009; Tickle et al., 2004). Ability to obtain stable and homogenous protein is important during screening of crystallization conditions. Complementary biophysical methods, such as thermal shift assays and analytical gel-filtration, are helpful in determining optimal constructs that are suitable for crystallization (Holcomb et al., 2017; Stevens, 2000). Using thermal shift assays, different buffer conditions that are most suitable for protein stability could also be screened. Various approaches in improving the diffraction resolution have been explored and systematic reviews on these approaches are available for optimization of the crystals (Zheng et al., 2014). The beamlines at most synchrotrons have improved so much over the past years that it now only takes a few minutes to collect a complete data set. The sample handling robots,

availability of remote data collection and microfocus beamlines are also new developments in synchrotrons that made the data collection a convenient process (Shi, 2014; Zheng et al., 2015). One of the most important developments that lead to determination of numerous structures is the availability of GUI version of the software used for the data reduction, validation and refinement (Adams et al., 2010; Knapman, 2001; Kuhn et al., 2002a; Potterton et al., 2003). These GUI versions of the crystallographic software suites have enabled the researchers to use these program suites without any knowledge of programming. Currently the most frequently used data reduction softwares are HKL-3000, XDS and iMOSFLM (Minor et al., 2006; Potterton et al., 2003). These automated programs provide easy visualization of collected images and integration of the diffraction spots. CCP4 suites provide a range of programs to choose from scaling the integrated data to molecular replacement, refinement and model analysis (Potterton et al., 2003). Another program package, PHENIX GUI provides a simple and powerful framework for data validation, ligand building, molecular replacement and model refinement (Adams et al., 2010). Along with PHENIX, visualization of electron densities and manual refinement can be performed using COOT (Emsley and Cowtan, 2004). Via COOT, chemical restraints of peptides such as Ramachandran outliers could be fixed manually (Emsley and Cowtan, 2004; Emsley et al., 2010). These automated softwares have allowed even novice structural biologists to be able to determine many macromolecular structures. Nonetheless, to be able to utilize these programs suites at the optimum, understanding the background theory of crystallography is necessary (Zheng et al., 2015).

Due to the above mentioned advancements, X-ray crystallography has become increasingly useful in drug-discovery in both screening stage and optimization stage. One example of the application of X-ray crystallography in structure based drug design was aliskiren, an inhibitor of renin. The renin is an aspartic protease that is important in regulation of blood pressure.

The inhibitors of renin are used in treatment for hypertension. Aliskiren, a non-peptidic renin inhibitor was discovered by analyzing the crystal structures of renin-inhibitor complexes and subsequently applying computational methods to optimize the inhibitor, aliskiren (Rahuel et al., 2000; Wood et al., 2003). It advanced to clinical trials and was approved by FDA for market in 2007(Gradman et al., 2005; Staessen et al., 2006).

Figure 1.21. Work flow of structure determination by X-ray crystallography. Figure adapted from (Abola et al., 2000)

From diffraction to 3D model

Protein crystals are 3D arrays of protein complexes. The assembly of these macromolecules in the crystal could be simplified using the crystal lattices system. The smallest derivative of

a crystal is a unit cell, on which n number of symmetry operations could be applied to achieve the whole crystal. The unit cells are denoted by its axes, a , b and c and the angles between them (α , β and γ). When the crystal is placed under the high energy monochromatic X-ray beam, the scattered beams which satisfy Bragg's law of diffraction gives a diffraction pattern on the detector (Figure 1.22).

Figure 1.22. Bragg's law of diffraction.

The collected diffraction patterns reflect the reciprocal lattices of the crystal which in turn reflects the real lattices of the crystal. For a given wavelength, for a certain angle of diffraction, the reciprocal lattice of the crystal will be found on an imaginary two dimensional circle called the Ewald's circle. Each diffraction pattern that is observed on the detector represents a spherical slice of the three dimensional reciprocal lattice. By rotating the crystal around, complete reciprocal lattice of the crystal could be recorded and reconstructed (Figure 1.23). The Fourier transformation of the reciprocal lattice gives to the real space lattice points of the crystal. Normally, during the crystal rotation, each lattice point in Miller indices is recorded more than once in consecutive diffraction images. These diffraction spots in different images belonging to the same spot in Miller index are merged and their intensities are normalized by scaling.

A

B

C

Figure 1.23. Ewald sphere construction. (A) The crystal at the centre of a sphere with radius $1/\lambda$ with an arbitrary origin of reciprocal lattice. (B) The lattice points that are on the Ewald sphere satisfy the Bragg diffraction equation and can be observed on the detector. (C) Turning of crystal around the same axis gives the diffraction spots of new reciprocal lattice points. S_0 represents the incoming X-ray beam. The A, B, C, D represent the points in reciprocal lattices. Green dots represent the lattice points that will be observed on the detector as diffraction spots. The grey dots represent the lattice points that will not be observed on the detector. Figure adapted from (Canduri and de Azevedo, 2008).

The diffraction of the protein crystals gives the diffraction data which represent the amplitudes of Fourier transform of the electron densities of the molecules (Abola et al., 2000). The 3D Fourier representation (structure factor) of these electron densities contains both amplitude and phase of diffracted waves (Evans, 2011; Kabsch, 2010a; Winn et al., 2011). The scaling and merging of diffraction intensities give the corresponding amplitudes of the structure factor but the information on phases is lost during the measurement (Adams et al., 2010; Evans, 2011; Kabsch, 2010a). The phase problem in X-ray crystallography is overcome in common by two approaches, (i) experimental phasing by anomalous X-ray scattering or multiple isomorphous replacement and (ii) by using molecular replacement (MR) where a related/known 3D model is used to estimate the phases (Dauter et al., 2002; Dodson, 2003; McCoy et al., 2007a). During the anomalous scattering, the phase information is obtained for either a single heavy atom derivative for single wavelength anomalous dispersion (SAD) or the methionine residues in the protein are labelled with selenium and their phase information is determined using multiple wavelength anomalous dispersion (MAD) (Dauter et al., 2002; Ealick, 2000). With many crystal structures deposited in PDB, the molecular replacement (MR) has become increasingly useful when the homologous protein structure is available. The homologous model is first rotated to be in the same

orientation with the target molecule in the reciprocal lattice (McCoy et al., 2007). Then the model is again translated throughout the unit cell (Mccoy et al., 2007). The MR method is mostly used when x-ray crystallography is applied in drug discovery for characterization of the detailed interactions between the hit compound and the target protein. The initial model output from the MR is inspected by the crystallographer and refined by both automatic programs and by manual improvements of the model (Adams et al., 2010; Afonine et al., 2012; Emsley and Cowtan, 2004; Emsley et al., 2010). The correctness of the model is judged by the difference between the observed diffraction amplitudes (F_o) and the amplitudes from the model (F_c). The chemical restraints based on the innate biochemical properties of peptides, such as bond lengths, dihedral angles, planar restraints and torsion restraints are used to further improved the model (Afonine et al., 2010a). Rounds of refinement and model improvements are carried out until the statistics of the model is acceptable. The coordinates of the model as well as structure factors of diffraction patterns are then deposited into the PDB for annotation.

Chapter 1.6: NS3 protease: An antiviral target

Section 1.6.1: General mechanism of proteases

Proteases, enzymes that hydrolyze the peptide bond, play an important role in many biological pathways, immune responses and infections and are found in every domain of life. The proteases can be divided into five classes depending on their catalytic mechanism: 1) metalloproteinases, 2) serine proteases 3) cysteine proteases 4) threonine proteases 5) aspartic acid proteases (Figure 1.24). The serine, cysteine and threonine protease have an active site residue that is involved directly during the catalysis while metalloproteinases and aspartic acid proteases hydrolyze the peptide bond via a highly reactive water molecule. The active site of serine, cysteine and threonine proteases contain an active nucleophile and a basic residue. All three proteases form a transition state called the tetrahedral intermediate. The oxyanion of the intermediate is stabilized by positive charge pocket formed by several hydrogen atoms of nearby residues and is called oxyanion hole. In the case of serine proteases, the oxyanion hole is formed by backbone NH groups of residues upstream of the catalytic serine.

Figure 1.24. Catalytic mechanism of (A) serine (B) cysteine (D) threonine and (D) metalloproteases. Figure adapted from (Erez et al., 2009)

Hydrolysis of a peptide by the serine protease involve three steps, (1) formation of enzyme-substrate intermediate (E-S) complex, followed by step (2) acylation of the serine and release of N-terminal peptide, and (3) release of C-terminal product by hydrolysis and subsequent regeneration of the serine/catalytic site (Figure 1.25). The acylation is represented by rate constant k_3 and deacylation by k_5 (Figure 1.25). Most biochemical characterization of the serine proteases are depicted in two macroscopic rate / steady state constants, k_{cat} and K_m which are derived from the equation provided in Figure 1.25B. Li et al., has shown that in DENV protease, the acylation is the rate limiting step (Li et al., 2005). In this case, the $k_5 \gg k_3$, and $k_{cat} = k_3$ and $K_m = K_d$ (Li et al., 2005).

Figure 1.25. Catalytic mechanisms of serine proteases. The rate constant k_3 represents the rate of acylation and k_5 the rate of deacylation. Overall rate constant equation for serine protease mechanism. For NS3 protease, the acylation is the rate limiting step. Figure adapted from Li et al., (Li et al., 2005).

Section 1.6.2: Structural biology of NS2B-NS3 protease

All genera of flavivirus harbour a serine protease as their non-structural protein 3 for the polyprotein processing. However, the role of the co-factor is slightly different between the hepaciviruses and flaviviruses. In HCV, the NS4A cofactor is involved in proper folding of the protease but does not interact with active site and is not involved in substrate binding (Hahm et al., 1995). For flavivirus, the NS2B cofactor wraps around NS3 protease and participates in substrate binding and catalysis. The HCV NS3/4A requires metal ion Zn^{2+} for the catalysis while it is not the case for the flavivirus NS2B/NS3 protease. NS3 protease from flavivirus adopts a trypsin like fold with the catalytic triad, His, Asp and Ser. NS3 protease requires the central hydrophilic cofactor region of NS2B for the catalytic activity (Clum et al., 1997). The N- and C-termini of NS2B are hydrophobic regions (Brinkworth et al., 1999). Although not involved in catalytic activity, these regions anchor NS3 protease near

the membrane (Droll et al., 2000). Mutagenesis studies of NS2B and NS3 protease domain have mapped amino acids that are important for protease activity. Early mutagenesis studies by Falgout et al., discovered that mutations at residues 129-135 and 148-153 either abolished or reduced the protease activity depending on the type of mutation (Valle and Falgout, 1998). In later mutagenesis studies, fluorogenic substrates are used to determine the effect of mutagenesis on protease activities (Chappell et al., 2005; Chappell et al., 2008; Niyomrattanakit et al., 2004; Shiryaev et al., 2007). In studies utilizing fluorogenic substrates, the mutations affect either K_m or k_{cat} or both (Chappell et al., 2005; Niyomrattanakit et al., 2004).

Crystal structures of NS2B-NS3 protease

X-ray crystallography has been a very useful tool in elucidating the detailed interactions of enzyme-substrate/inhibitor. The first NS3 protease structure was reported by Erbel et al., in 2006 using the recombinant NS2B-NS3 construct, where 47 residues of NS2B cofactor are tethered to NS3 protease by an artificial glycine linker. The subsequent NS2B-NS3 structures from flaviviruses used this construct design. So far, both free enzyme structures and enzyme-inhibitor complex structures have been reported for DENV and WNV NS3 proteases.

The free enzyme structures of DENV1, DENV2 and WNV NS3 proteases showed very similar overall fold of NS2B-NS3 protease. The NS3 protease forms two β barrels comprising six beta strands in each. The N-terminus of the NS2B cofactor contributes to the first β barrel by forming a β strand. The C-terminal residues of NS2B cofactor are flexible and not observed in the crystal structure (Aleshin et al., 2007; Chandramouli et al., 2010; Erbel et al., 2006). The free enzyme protease structure of ZIKV protease was reported recently and showed similar overall fold to WNV and DENV protease structures with the C-terminus of NS2B cofactor being flexible and missing from the structure. These free enzyme structures with flexible NS2B C-terminus are considered as the “open” conformation. When

NS2B-NS3 protease is in complex with an inhibitor, the NS2B C-terminal residues fold back into a β hairpin and contribute to the active site. This conformation of NS2B-NS3 protease is considered as the “closed” conformation.

The NS2B-NS3 protease in complex with peptidomimetic inhibitor structures have also been reported for WNV, DENV and ZIKV protease (Kang et al., 2017) (Figure 1.26). The crystal structures of NS2B-NS3 protease are listed in table 1.4. The first NS3 protease-inhibitor structure was reported in 2006 by Erbel et al., with benzoyl-norleucine-lys-arg-arg-H (Bz-nKRR-H) (Erbel et al., 2006). The structure highlighted the role of NS2B in substrate binding which is in contrast to HCV cofactor NS4A. The structure also revealed the residues involved substrate binding pockets of NS2B-NS3 protease. The S1-S3 pockets of the protease are well defined with NS2B cofactor contributing to S2 and S3 pockets (Erbel et al., 2006). The S4 pocket is found to be shallow, solvent exposed and hydrophobic. The tetrapeptide inhibitor is bound to the active site with a horse-shoe like shape backbone with P1-P3, Lys-Arg-Arg side chains filling up the S1-S3 sites (Erbel et al., 2006). The S1 pocket is lined by G151, Y161, Y150 and D129 side chains and backbones of Y130-T132 (Erbel et al., 2006). The P2 Arg side chain interacts with N152 and D75 of catalytic triad as well as with N82, G83 and N84 of NS2B. The P3-P4 residues are solvent exposed and not tightly bound to the active side. Four more protein peptidomimetic inhibitor complex structures were solved and these inhibitors have P1-P2, Lys-Arg or Arg-Arg with C- and N- terminal modifications (Hammamy et al., 2013; Lei et al., 2016; Noble et al., 2012; Robin et al., 2009). In all these structures, the horse shoe like back-bone formation of the substrate is conserved (Hammamy et al., 2013; Lei et al., 2016; Noble et al., 2012; Robin et al., 2009). Very little variations between P1-S1 and P2-S2 interactions are observed as well indicating that the P1-P2 interactions are conserved across the flavivirus genus. The newest protease-inhibitor structure was that of ZIKV NS2B-NS3 protease with a boronic inhibitor (Lei et al., 2016). The

structure showed the boronic moiety forming a covalent bond with the catalytic serine. The P1 and P2 substrates occupy the pockets similar to that of WNV and DENV protease. Although these inhibitors showed different P4 cap moieties, the aromatic ring of the P4 cap is facing in parallel manner to the active site and most of the cases, the P4 backbone interacts with P1 via electrostatic interactions. This could be a way to maintain the horseshoe like backbone structure of the inhibitor. The bovine pancreatic trypsin inhibitor (BPTI), a strong basic protein of about 6 kDa, is a potent inhibitor of NS2B-NS3 protease with K_i value in low/sub nanomolar range. Both DENV and WNV protease in complex with BPTI have been reported (Aleshin et al., 2007; Noble et al., 2012). Only WNV protease in complex with BPTI showed the NS2B cofactor C-terminus assuming the beta hairpin loop while in DENV, the C terminus of the cofactor was folded backwards and form a beta sheet with neighboring NS2B C-terminus as a result of crystal artefact (Aleshin et al., 2007; Noble et al., 2012). Again in WNV-BPTI structure, the backbone of the BPTI could be superimposed well onto those of Peptidomimetic inhibitors. The S1 and S2 pockets are occupied by C14 and K15 of BPTI in the similar manner as Arg and Lys in peptidomimetic inhibitors (Figure 1.26).

Figure 1.26 Crystal structure of NS2B-NS3 protease with peptidomimetic inhibitors. (A) Superimposition of peptidomimetic inhibitors in the binding pocket of NS2B-NS3 protease. S4-S1' pockets are labelled. Inhibitors are shown in sticks, 2-naphthoyl-KKR-H in yellow (PDB id: 3E90), benzoyl-NKRR-H in light grey (PDB id: 2FP7) and in orange (PDB id: 3U1I), Cn-716 in cyan (PDB id: 5LC0), 3,4-dichlorophenylacetyl-KK-GCMA in magenta (PDB id: 2YOL). (B) Residues from NS2B cofactor and NS3 protease that are involved in substrate binding pockets. The NS2B residues are labelled in magenta with residue number and one letter code, and NS3 residues are labelled in black in the same manner. The residues shown are from WNV NS2B-NS3 protease structure (in magenta PDB id: 2YOL), DENV NS2B-NS3 protease structure (in orange PDB id: 3U1I) and ZIKV NS2B-NS3 protease structure (light grey PDB id: 5LC0).

Table 1.4. List of flavivirus NS2BNS3 protease structures

Virus	PDB code	Ligands	Remarks	References
DENV1	3L6P	-	Open	(Chandramouli et al., 2010)
	3LKW	-	open	
DENV2	2FOM	-	open	(Erbel et al., 2006)
	4M9F		open	(Yildiz et al., 2013)
	4M9I		open	
	4M9K		open	
	4M9M		open	
	4M9T	DTNB, a covalent allosteric inhibitor	open	
DENV3	3U1I	Bz-Nle-Lys-Arg-Arg-H	closed	(Noble et al., 2012)
	3U1J	Aprotinin	open	
JEV			open (NO NS2B c terminal)	(Weinert et al., 2015)
	4R8T	-		
WNV	2YOL	3,4-dichlorophenylacetyl-Lys-Lys-GCMA	closed	(Hammamy M Z, et al., 2013)
	3E90	2-naphthoyl-Lys-Lys-Arg-H	closed	(Robin et al., 2009)
	2FP7	Bz-Nle-Lys-Arg-Arg-H	closed	(Erbel et al., 2006)
	2IJO	BPTI	closed	(Aleshin et al., 2007)
	2GGV	-	open	
	5IDK	Bz-(4-CH ₂ NH ₂)-Phe-Arg-H	closed	(Nitsche et al., 2017)
ZIKV	5LC0	Cn176	closed	Lei et al., 2016
	5T1V		open	(Lee et al., 2016)
	5GXJ		open	(Chen et al., 2016)

NMR studies of NS3 protease

The NMR studies of NS2B-NS3 protease provide the dynamics of protease domain in complement to x-ray crystallography. Early NMR studies of both WNV and DENV protease used the conventional glycine linked construct utilized in X-ray crystallography. The ^1H - ^{15}N HSQC spectra of these constructs are not well dispersed suggesting the presence of more than one conformation (de la Cruz et al., 2011; Su et al., 2009a). Titration of NMR samples with inhibitor resulted in well dispersed spectra for these constructs showing the stabilization of the protein by inhibitor (de la Cruz et al., 2011). It was Kim et al., who first showed that well dispersed cross peaks are obtained when there is no covalent linkage between NS2B cofactor and NS3 protease for DENV (Kim et al., 2013) (Figure 1.27). The NMR spectrum of the unlinked free NS2B-NS3 protease is similar to glycine linked NS2B-NS3 protease-inhibitor spectra proposing that unlinked NS2B-NS3 protease in solution is predominantly in closed conformation (de la Cruz et al., 2014; Kim et al., 2013). They also demonstrated that the presence of the glycine linker between the NS2B and the NS3 protease promoted the conformation exchange of NS2B (de la Cruz et al., 2014) (Figure 1.27).

Figure 1.27. NMR studies of NS2B-NS3 protease. (A) ^1H - ^{15}N HSQC spectrum of DENV2 NS2B-NS3 protease with artificial glycine linker. The peaks are not well dispersed making the residues assignment challenging. (B) ^1H - ^{15}N HSQC spectrum of unlinked DENV2 NS2B-NS3 protease showing dispersed spectra. (C) ^1H - ^{15}N HSQC spectra of ^{15}N -Ile-labelled DENV2 unlinked NS2B-NS3 protease and (D) glycine linked NS2B-NS3 protease and (E) of glycine linked NS2B-NS3 protease in complex with a generic serine protease inhibitor (p-Nitrophenyl-p'-guanidinobenzoate HCl). The spectra of unlinked NS2B-NS3 protease (C) resembles the spectra of linked NS2B-NS3 protease-inhibitor complex (E). Figures adapted from (de la Cruz et al., 2014; Kim et al., 2013)

Specificity of flavivirus proteases

Since the past two decades, studies establishing the specificity of flavivirus proteases especially for YFV, WNV, and DENV have been carried out, and recently for ZIKV protease (Bessaud et al., 2006; Chanprapaph et al., 2005; Chappell et al., 2005; Chappell et al., 2006; Chen et al., 2016; Lee et al., 2016; Leung et al., 2001; Li et al., 2005; Nitsche et al., 2017; Niyomrattanakit et al., 2005; Yusof et al., 2000). The studies using crystallography and molecular simulation along with biochemical assays and mutagenesis studies have mapped the substrate binding pockets of NS2B-NS3 protease as well as their substrate preferences (Chanprapaph et al., 2005; Chappell et al., 2005; Chen et al., 2016; Junaid et al., 2012; Lee et al., 2016; Leung et al., 2001; Li et al., 2005; Li et al., 2017a; Lohr et al., 2007; Mueller et al., 2007; Nitsche et al., 2017; Niyomrattanakit et al., 2005; Noble et al., 2012; Prusis et al., 2008; Schuller et al., 2011; Yusof et al., 2000). The polyprotein cleavage sites of NS2B-NS3 protease for different flavivirus species are shown in table 1.2. In general, the NS3 protease recognizes the dibasic residues at P1 and P2 position with small residues at P1' and P2' positions (Leung et al., 2001; Li et al., 2005; Yusof et al., 2000). The substrate specificity of flavivirus DENV, WNV, YFV and ZIKV proteases rely heavily on the P2-P2' interactions with only either Lys or Arg at P1 or P2 sites are tolerated while variations of even bulky amino acids in P3 and P4 positions are allowed (Bera et al., 2007; Gruba et al., 2016; Li et al., 2005; Mueller et al., 2007). For P2 residues, DENV NS3 protease prefers P2 Arg while WNV JEV and MVEV protease prefers P2 Lys (Li et al., 2005). Shiryaev et al., also showed that WNV protease does not tolerate longer, hydrophobic residues at P1' and P2' while DENV protease does (Shiryaev et al., 2007). The protease substrate specificity is an interdependent mechanism with variations in extended substrate sites contributing to both catalysis rate and binding of inhibitor. Therefore, it is still challenging to predict the substrate

specificity of these proteases for design or optimization of peptide inhibitors despite the extensive studies on substrate specificity.

Section 1.6.3: Antivirals targeting NS2B-NS3 protease

Although HCV NS3/NS4A protease inhibitors have been approved as antiviral drugs since 2011, the search for the inhibitors against flaviviral protease is still in the stage of lead finding and optimization (Low et al., 2017). The substrate binding site of the protease is shallow and solvent exposed with a strong preference for multi-basic inhibitors. Many studies have been reported the development of the peptide inhibitors with modifications. These substrate mimics usually have either (1) C-terminal electrophilic group or (2) N-terminal cap or (3) non-natural amino acids at P1 and P2 positions.

The first peptide inhibitors of DENV protease are characterized in 2001 by Leung et al., using *p*-nitroanilide substrates (Leung et al., 2001). The peptide inhibitor based on NS3/4A cleavage sequence ($^{\text{P6}}\text{FAAGRRSL}^{\text{P2'}}$) with an α ketoamide at the C-terminus showed a lowered K_i compared to the inhibitor with NS4B/5 cleavage sequence ($^{\text{P6}}\text{TTSTRRSL}^{\text{P2'}}$) with the same α keto group (Leung et al., 2001). Replacing the α -keto group of the inhibitor with an aldehyde warhead and removal of P1' and P2' residues further decreased the K_i to 16 μM (Leung et al., 2001). More peptidic inhibitors were characterized in the study by Chanprapaph et al., with tripeptide inhibitors showing micro molar K_i against DENV NS2B-NS3 protease (Chanprapaph et al., 2005). Based on the results of Chanprapaph et al, Yin et al., investigated the effect of C-terminal electrophilic groups on benzoyl-norleucine-lysine-arginine-arginine-X inhibitors (Bz-nKRR-X) (Yin et al., 2006b). They found that addition of aldehyde moiety at the C-terminus of peptide inhibitor reduced the K_i by 100 fold compared to the carboxylic acid (Yin et al., 2006b). Other two C-terminal electrophile groups that increased the potency are trifluoromethylketone and boronic acid (Yin et al., 2006b). In the second part of their study, Yin et al., found that with the aldehyde group at the C-terminus,

P1 Arg could be replaced by Phe and Trp residues with little changes in K_i (Yin et al., 2006a; Yin et al., 2006b). Although the *in vitro* inhibition efficiencies of these peptidic inhibitors are in low micromolar to nanomolar range, both aldehydes and trifluoromethylketone also bind to the human proteases raising selectivity issues (Nitsche et al., 2017). Recently, Nitsche et al., further characterized the peptidomimetic compounds with boronic acid at the C-terminus across DENV, WNV and ZIKV proteases *in vitro* and as well as in cell culture (Nitsche et al., 2017). The study revealed a few interesting compounds with basic-lipophilic residues at P2 position with EC_{50} s in micromolar range and with cellular cytotoxicity (CC_{50}) higher than 100 μ M (Nitsche et al., 2017).

Using Bz-nKRR-H as the scaffold, the substrate profiling at P1-P4 positions were carried out (Li et al., 2005). The replacement of non-basic residues at P1-P3 position usually reduced the affinity, although peptide aldehydes with 4-methylphenylalanine or tryptophan at P1 residue have similar range of affinity to that of Bz-nKRR-H (Li et al., 2005). To increase the cellular permeability and pharmacokinetics of the inhibitors, P1-P2 basic residues were replaced with more lipophilic amino acids or un-natural peptides. Schüller et al., explored different N-terminal cap moieties using two scaffolds, X-KRR-H and X-KKR-H against both DENV and WNV protease (Schuller et al., 2011). No significant improvement was observed for N-terminal caps that were tested in the X-KRR-H series while 4-phenylphenylacetyl had the lowest K_i for the N-terminal caps tested in the X-KKR-H series for DENV protease. The significant influence of cap moiety on K_i was demonstrated by loss of inhibition when N-terminus was changed from 2-phenylacetyl- to 3-phenylpropanoyl- group for DENV protease (Schuller et al., 2011). In the case of WNV protease, 4-phenylphenylacetyl cap at N-terminus reduced IC_{50} by 10 folds for X-KRR-H (Schuller et al., 2011). The study also showed that for DENV protease, Arg at P2 position is preferred while for WNV; Lys at P2 position is preferred (Schuller et al., 2011). Tripeptide agmatine inhibitors with different N-terminal

caps were screened for WNV protease and found that inhibitors with phenylacetyl N-terminal cap showed the lowest IC_{50} s (Lim et al., 2013). In another study by Hammamy et al., the P1 residue was replaced by 4-guanidino cyclohexylamide with a phenylacetyl as an N-cap (Hammamy et al., 2013). This inhibitor was further optimized by varying the P2, P3 and P4 positions (Kouretova et al., 2017). Despite their good *in vitro* inhibition potency, due to their polar nature, these compounds did not show any inhibition in cell based assays (Kouretova et al., 2017). Weigel et al., characterized the P1 and P2 positions of similar hybrid peptidic inhibitors (Weigel et al., 2015). In the study, an inhibitor with phenylpropanoic acid at P1 position and 4-amidino-phenylalanine at P2 has an IC_{50} of 0.25 μ M against DENV2 protease with no detectable inhibition on trypsin (Weigel et al., 2015). The same inhibitor inhibited the viral replication by 58.8% at 50 μ M concentration in plaque assay (Weigel et al., 2015). Dipeptide inhibitors in the form of methionine-proline anilides characterized by Zhou et al., showed K_i value of less than 20 μ M (Zhou et al., 2013). Nitsche et al., developed small retro peptide inhibitors by fusing small molecular synthetic scaffolds with peptide sequence in the manner of **Cap**-RKN-H in the attempt to remove the electrophilic warhead from C-terminus (Nitsche et al., 2012) and the optimal cap was found to be arylcyanoacrylamide group. These results were followed by more modifications at the cap moiety, leading to rhodamine- and thiazolidinedione-based peptides (Nitsche et al., 2013). These compounds showed antiviral activity in cell culture assays with EC_{50} values in the micro molar range (Nitsche et al., 2013). Different C-terminus moieties for benzoyl-tripeptides were also investigated by Behnam et al., and Weigel et al., (Behnam et al., 2014; Weigel et al., 2015). In another study by Behnam et al., tripeptides with 4-hydroxyphenylglycine as P1 residues were optimized at both N- and C- termini (Behnam et al., 2015). The optimized peptide has an EC_{50} of 15.5 μ M with $CC_{50} > 100$ μ M (Behnam et al., 2015). Fragment merging of two peptidic inhibitors, one with a thiazolidinone C-terminal cap and another with phenylglycine at P1 position

resulted in a new hybrid peptidic inhibitor with 20 fold higher potency (Behnam et al., 2014). Starting with octapeptide substrate based peptidic inhibitors (RRRRXXXX-NH₂), Prusis et al., successively removed the arginines which give the non-substrate based tetrapeptide, WYCW-NH₂ with K_i less than 5 μ M (Prusis et al., 2013). Studies involving cyclic peptides as inhibitors were also reported. The recombinant cyclic peptide, retrocyclin-1, inhibits viral replication in cell culture (Tomlinson and Watowich, 2012). Although their inhibitory properties are good, most of them showed cytotoxicity at relatively low concentration, and are therefore unlikely to be developed as drugs. Longer peptides such as plectasin (40 amino acids) showed non-competitive binding with K_i of 5.03 μ M (Rothan et al., 2013) but were cytotoxic at concentration as low as 20 μ M. Table 1.5 summarizes the different peptidomimetic inhibitors with IC₅₀s or EC₅₀s in low micromolar range with CC₅₀s higher than 100 μ M.

Table 1.5. Peptidomimetic inhibitors of NS2B-NS3 protease

No.	Structure	IC ₅₀ (μ M)	EC ₅₀ (μ M)	CC ₅₀ (μ M)	Reference
1		0.036 (DENV) 0.071 (WNV) 0.83 (ZKV)	18 (WNV)	>100	(Nitsche et al., 2017)

2		6.1 (DENV)	16.7	>100	(Nitsche et al., 2013)
3			15.5	>100	(Behnam et al., 2015)

Another approach towards flaviviral protease inhibitor is non-peptidic compounds. Owing to its shallow solvent exposed pocket, the NS3 protease is challenging for the development of small molecule inhibitors. Even irreversible protease inhibitors such as (pheynyl-methylsulfonyl fluoride (PMSF) have no inhibitory activity against NS3 protease (Leung et al., 2001). Small molecule inhibitors are usually discovered via high throughput screening of compound libraries and virtual screening approaches. Although crystals structures of N2BNS3 protease in complex with peptidic inhibitors have been determined, no crystal structure of NS3 protease in complex with a small molecule inhibitor is available making the structure based drug design challenging. Early stages of small molecule compounds used a guanidino group to mimic the P1 arginine and two of these compounds showed IC₅₀s of 14 and 23 μ M (Ganesh et al., 2005). A guanidine-substituted phenyl benzoates compound was identified during later virtual screening and was found to have a K_i of 2 μ M (Knehans et al., 2011). Phenolic and polyphenolic compounds were also identified during a high throughput screen (Tomlinson et al., 2009). Further studies done on these compounds showed high cytotoxicity (Rothan et al., 2014). Quinoline and hydroxyquinoline based compounds were

identified during HTS screening with $K_i=17\ \mu\text{M}$ for hydroxyquinoline and $\text{IC}_{50}=9.45\ \mu\text{M}$ (Deng et al., 2012; Lai et al., 2013). One of the quinoline derivative compounds has EC_{50} of $24.7\ \mu\text{M}$ with low cytotoxicity ($\text{CC}_{50} < 100\ \mu\text{M}$) (Deng et al., 2012). The quinoline moiety is expected to form interactions with Y160 via π stacking. Another HTS study identified three compounds against as DENV protease inhibitors with K_i of less than $5\ \mu\text{M}$ but no cell based assay activity was available (Nguyen et al., 2013). Non-competitive small molecule inhibitors were also identified during virtual screening campaign by a few groups (Li et al., 2017c; Shiryayev et al., 2011). These compounds in general target NS2B cofactor binding to NS3 protease. Shiryayev et al., identified two structurally different compounds that are active against WNV and DENV with IC_{50} values less than $5\ \mu\text{M}$ (Shiryayev et al., 2011). These compounds also have antiviral activity in cell based assays (EC_{50} of $39\ \mu\text{M}$ and $59\ \mu\text{M}$) as well as low cytotoxicity values (CC_{50} of $213\ \mu\text{M}$ and $117\ \mu\text{M}$) (Shiryayev et al., 2011). Pambudi et al. also identified and validated a small molecule compound targeting the NS2B binding site (Pambudi et al., 2013). This compound has high antiviral activity across all four serotypes of DENV with EC_{50} less than $2.5\ \mu\text{M}$ with low toxicity ($\text{CC}_{50} = 67.3\ \mu\text{M}$). However, the same compound is less effective for other flaviviruses such as JEV ($\text{EC}_{50} = 29.8\ \mu\text{M}$) demonstrating the selectivity towards DENV.

The allosteric inhibitors of NS2B-NS3 protease offer an alternative route to bypass the issue of targeting the challenging shallow pocket of NS2B-NS3 protease. Li et al., in 2017 identified highly promising allosteric inhibitors that are approved existing drugs via a novel complement FRET assay (Li et al., 2017c). Three hit drugs, Nitazoxanide, Niclosamide and Temoporfin, inhibited DENV protease with IC_{50} s ranging from 1.1 - $15.9\ \mu\text{M}$. These compounds showed good antiviral activity against flaviviruses in cell based assays (Li et al., 2017c). The Temoporfin and Nitazoxanide had CC_{50} of 40.7 and $77\ \mu\text{M}$ indicating moderate cytotoxicity. But Niclosamide had high cytotoxicity with CC_{50} of $4.8\ \mu\text{M}$. Reduced viremia

and 83% survival rates were observed when A129 mice are treated with 1mg/kg Temoporfin for 5 days (Li et al., 2017c) (Table 1.6).

Table 1.6. Non-peptidic NS2B-NS3 protease inhibitors

K_i		1.1		9.45	
EC₅₀ μM	>50			24.7	
CC₅₀ μM	>100	77		>100	
References	(Steuer et al., 2011)	(Li et al., 2017c)		(Deng et al., 2012)

K_i		2.36		2.75	
EC₅₀ μM	0.74			39.9	
CC₅₀ μM	67.3			213	
References	(Pambudi et al., 2013)	(Ezgimen et al., 2012)		(Shiryaev et al., 2011)	

Chapter 1.7: Aim of the thesis

Due to the global warming and urbanization, the flavivirus infections are on the rise. DENV alone causes 390 million infections per year with the global health burden of about 39 million US dollars. The social and economic cost of the ZIKV outbreak in Latin Americas during 2015-2017 is estimated to be at US\$ 7-18 billion. Currently there is neither specific nor broad spectrum antiviral against the flavivirus infections available. In this study, we employ different biochemical and biophysical methods to characterize one of the most attractive drug-target proteins, NS3, from flavivirus. Previous NMR studies suggested that different construct designs of NS2B-NS3 lead to different conformational dynamics of the protease. Prior structural studies of flavivirus NS3 protease and full length are solely based on the construct where NS2B is tethered to NS3 with an artificial glycine linker. In addition, only a limited numbers of NS3 protease-inhibitor complex crystal structures are available. All the available protease-inhibitor complex structures are based on peptidomimetics inhibitors. The biochemical and structural studies of full length NS3 with focus on the protease domain is not yet available. Therefore, in this thesis, we report the characterization of structural and biochemical properties of the full length NS3 and NS3 protease via the following aims.

Specific aims:

- 1) Structural and enzymatic activity characterization of different constructs of NS3 protease from DENV and ZIKV virus to understand the effect of the linker between NS2B-NS3 on structural and biochemical properties of NS3 protease.
- 2) Structural characterization of NS3 protease in complex with a) peptidomimetics inhibitors and b) small molecule inhibitors to provide structural bases of NS3 protease in recognizing different substrates.

- 3) Structural and enzymatic activity characterization of different constructs of NS3 full length from DENV to understand the effect of linker between NS2B cofactor and NS3 protease on domain conformations and biochemical activities of full length NS3.

2. Chapter 2: Materials and methods

Chapter 2.1: Molecular cloning and site directed mutagenesis

DENV NS3 full length plasmids

The recombinant plasmid containing wild type NS3 linked to co factor NS2B residues 49—96 was received from Prof. Subhash Vasudevan Lab (SV Lab) along with two mutants of same plasmid S135A and H51N. Hence, two mutants, each with a mutated residue in one of the catalytic triads were created by SV lab. All other mutants were created by site directed mutagenesis (SDM) at DL lab. The gene bank accession number is JF262783.1.

DENV NS3 protease plasmids

DENV4 NS3 protease plasmids are generated by SDM by introducing a STOP codon at the end of the protease domain. The bivalent constructs for DENV are ordered from Bio Basic Asia Pte Ltd, Singapore. The gene bank accession number is JF262783.1.

ZIKV protease plasmids

The ZIKV protease plasmids, bZiPro, eZiPro and gZiPro were received from Dr. Congbao Kang Lab, Experimental therapeutics centre (ETC), A*STAR. The mutagenesis on bZiPro was performed by SDM. The gene bank accession number is AEN75265.1.

For SDM, the primers were designed using ApE. plasmid editor. and ordered from Integrated DNA technologies, Singapore (IDT.sg). The primers were diluted to 10 μ M in sterilized water. The PCR reaction was set up using Phusion Flash High-Fidelity PCR MasterMix or KOD HotStart MasterMix. The 30 μ L reaction consists of 50 - 100 ng template plasmid, 1 μ L of forward and reverse primers at 10 μ M concentrations, and 15 μ L of 2xMastermix. The template strands were separated by heating to 95°C for 3-5 minutes, followed by 20-30 times repeat of PCR amplification cycle. The cycle consists of 30 second denaturation at 95°C, 55

second annealing at (T_m of primer - 5)°C and extension at 45 seconds per kilo base pair at 72°C. The final extension step was introduced for 10 minutes at 72°C. The template plasmids were then digested by adding 0.5 µL of DpnI enzyme (NewEngland biolabs) and incubated for overnight at 37°C. The reaction was transformed into transformation competent *Escherichia coli*. The transformants were recovered by addition of 400 µL of LB and incubating at 37°C for two hours. The transformants were plated on LB agar plate supplemented with respective antibiotics and incubated at 37 °C overnight. The positive colonies were picked the next day and subsequently grown in 4 ml of LB with respective antibiotic. Plasmids were exacted the following day and mutagenesis was confirmed by sequencing.

Chapter 2.2: Expression and purification

The plasmids were transformed into *E.coli* BL21 (DE3). Transformants were grown in Luria Broth (LB) medium supplemented with ampicillin (100 mg/L) or kanamycin (50 mg/L) and chloramphenicol (37 mg/L) at 37°C until OD₆₀₀ of 0.8-1.0 was reached. The culture was cooled to 18°C and the overexpression of the recombinant protein was subsequently induced by addition of 1mM IPTG. Cells were harvested after 15 hours by centrifugation at 5000 rpm for 20 minutes at 4°C. Cells were resuspended in lysis buffer (25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 8.0, 500 mM sodium chloride (NaCl), 5 mM β-mercaptoethanol (β-ME), 5% glycerol, 10 mM imidazole) at a ratio of 20-25 ml of lysis buffer per 5 mg of cell pallet. Cells were lysed with homogenizer (GEA Niro Soavi) at 900 bar for 5-10 minutes. The soluble fraction was obtained by centrifugation of lysate at 40,000 rpm for 30 minutes at 4°C. The soluble fraction was collected; Ni-NTA Agarose beads (Qiagen) equilibrated with lysis buffer was added to the fraction and incubated for 2 hours at 4°C. The recombinant proteins bound to Ni-NTA beads were retained on the column bed. The beads were then washed extensively with lysis buffer containing 20-40 mM imidazole;

bound proteins were eluted with lysis buffer containing 300 mM imidazole. N terminal histidine-tag was cleaved by Tobacco Etch Virus (TEV) protease during the dialysis of the eluted fraction against size exclusion chromatography (SEC) buffer (25mM HEPES, pH 7.5, 150 mM NaCl, 2 mM DTT, 5% glycerol) overnight. The TEV protease was added with 1:40 ratio of TEV protease: target protein. For ZIKV protease constructs, the His-tag was cleaved by addition of thrombin protease in the same dialysis procedure. The thrombin protease was purchased from GE Healthcare (Research Instruments Pte. Ltd, Singapore). His-tag and the TEV protease/ the thrombin protease were removed by running through Ni-NTA Agarose beads equilibrated with SEC buffer (25mM HEPES, pH 7.5, 150 mM NaCl, 2mM DTT, 5% glycerol) and flow through fraction was collected. For DENV full length constructs, the His-tag cleaved proteins were polished by running through HiTrap Heparin HP 5 ml column (GE Healthcare) equilibrated with Heparin buffer (25mM HEPES, pH 7.5, 2mM DTT, 5% glycerol) with 100mM NaCl. After washing extensively with Heparin Buffer with 150 mM NaCl, the proteins were eluted at Heparin buffer containing 300 mM NaCl. Fractions were pooled, concentrated and finally applied to HiLoad 16/600 Superdex 200 pg (GE Healthcare) column for size exclusion chromatography. For the protease constructs, the Heparin step is omitted with His-tag cleaved proteins directly injected to HiLoad 16/600 Superdex 75 pg column for size exclusion chromatography. Fractions were collected, concentrated using Vivaspin 20 with molecular weight (MW) cut off 30 kDa from Sartorius Singapore Pte Ltd, Singapore for full length constructs and using Vivaspin 10 with MW cutoff 10 kDa by centrifugation at 4000g and flash frozen with liquid nitrogen in small aliquots for further experiments.

Chapter 2.3: Crystallization

Section 2.3.1: Crystal screening

For screening of crystallization conditions, we use crystallization kits from Hampton Research, California, USA (PEGIon Screen, GridScreen, CrystalScreen, and PEGRxScreen) and from Molecular Dimensions, UK (MIDAS, Proplex Screen). All HTS crystallization screenings were carried out by sitting drop vapor diffusion method. The protein and the crystallization solutions are mixed in 1:1 ratio by liquid handling robots, Phoenix or Mosquito. The INTELLI-PLATE® 96 (ArtRobbinsInstruments) were used. The plates were incubated at Rock Maker at 18 °C and the progress was monitored by capturing the images using Rock Imager 1000 at 0 day, 1 day, 3 days, 5 days, 2 weeks, one month and three months.

Section 2.3.2: Crystal optimization

Figure 2.1 Crystal optimization of different constructs. (A-B) gNS2B₄₇NS3 full length (C-D) eNS2B₄₇NS3 Pro, (E) bZiPro.

There are in general three groups of crystals, DENV NS3 full length crystals, DENV4 NS3 protease crystals, and ZIKV NS3 protease crystals. The crystal hits and optimized crystals are shown in Figure 2.1. The DENV full length proteins crystallized at similar crystallization conditions with pH ranging from 6.4-7.4 and PEG concentration of 8-12%. The detailed crystallization conditions for different proteins are given in the respective chapters where the results were discussed more in details. The DENV4 NS3 protease crystals are optimized by varying pH and further improved by dehydration. The ZIKV protease crystals grew at various crystallization conditions containing ammonium sulfate.

Chapter 2.4: Data collection, Structure determination and refinement

Diffraction intensities were recorded at PIXAIII at Swiss Light Source, MXI and MXII at Australian Synchrotron, Proxima beamline at French National Synchrotron Facility and TPS 05A at National Synchrotron Radiation Research Centre, Taiwan. Diffraction intensities were integrated using iMOSFLM, XDS or HKL 3000 (Battye et al., 2011; Kabsch, 2010a, b; Powell et al., 2007). Scaling and merging of the intensities were done using SCALA or AIMLESS from CCP4 suite (Diederichs and Karplus, 1997; Evans, 2007, 2011). The

qualities of the data sets are assessed by POINTLESS in CCP4 suite and Xtriage in PHENIX (Evans, 2011; Winn et al., 2011) (Adams et al., 2010). Data collection and refinement statistics for the crystals are stated in the respective chapters.

Section 2.4.1: Molecular replacement for DENV4 NS3 full length

Solution for gNS2B₄₇NS3 S135A conformation I and gNS2B₄₇NS3 S135A with BPTI were solved using PHASER MR (CCP4 suite) using DENV4 NS2B₁₈NS3 full length (PDB id: 2VBC) as the search model (Luo et al., 2008a; McCoy, 2007; McCoy et al., 2007a). Prior to molecular replacement, the solvents and ligands molecules were removed from the search model. For the molecular replacement of subsequent DENV NS3 full length data sets, gNS2B₄₇NS3 S135A free enzyme structure is separated into two coordinate files, one for protease domain and another for helicase domain. The structure solutions for the two domains were searched separately, with the helicase domain being searched first followed by the protease domain using PHASER MR in PHENIX (Bunkoczi et al., 2013; McCoy, 2007). The structure solutions were refined using PHENIX refine program (Afonine et al., 2012; Headd et al., 2012). The refined structures were checked and manually refined using Coot (Adams et al., 2010; Afonine et al., 2010b; Emsley et al., 2010). Solution for NS2B₄₇-NS3 S135A conformation II was found using NS2B₄₇-NS3 conformation I as search model and refined as described above.

Section 2.4.2: Molecular replacement for ZIKV protease structures

For the ZIKV protease crystals (eZiPro), after indexing the datasets were found to be twinned. The twinned data was analyzed and twin law was determined with Xtriage program on PHENIX (Adams et al., 2010). Due to the twinning, the data were indexed in apparent hexagonal space group (P6₂₂₂). The data set was reindexed from P3₁ using iMOSFLM (Battye et al., 2011). The molecular replacement for eZiPro was performed using the previous WNV protease structure after removing the bound inhibitor from the coordinate file. (PDB id: 2FP7). The models were then subjected to rounds of manual refinement using WinCoot and further refined using PHENIX refine (Afonine et al., 2012). The twin law (−h, −k, l) was used throughout the refinement process. For bZiPro, the data could be indexed into

two different unit cell dimensions. Some of the bZiPro datasets were twinned and some were not. The twin laws for the data sets were determined using Xtriage. The twin law was used during the refinement with Phenix.refine (Afonine et al., 2012). The molecular structures of the compounds were sketched using PubChem sketcher and saved as .mol files. The chemical restraints files and coordinate files of the compounds are built using Prodrug server (Schuttelkopf and van Aalten, 2004). The compounds were added using the findligand function in Coot (Emsley and Cowtan, 2004). The electron density maps were generated using Maps in PHENIX. The figures were prepared using Pymol (PyMOL The PyMOL Molecular Graphics System). Rotational and translational movements of domains were carried out using DynDom (CCP4 suites) and Superpose (CCP4 suites). Buried surface area between domains are calculated using PISA (EMBL) (Hayward and Berendsen, 1998; Hayward and Lee, 2002; Krissinel and Henrick, 2004, 2007).

Chapter 2.5: Biophysical and Biochemical assays

Section 2.5.1: Protease activity assay

The protease assays were carried out using peptide substrate with a fluorophore, Benzoyl-Nle-Lys-Arg-Arg-Aminomethylcoumarin (Bz-nKRR-AMC) (Peptide Institute, Japan). The protease cleaves the substrate releasing the fluorophore, aminomethyl coumarin (AMC). The AMC could be detected by excitation at 380 nm and emission at 450 nm (Figure. 2.2). The assay buffer used is 20 mM Trisaminomethane-hydrochloric acid (Tris-HCL) pH 8.5, 2 mM DTT, 10% glycerol and 0.01% Triton X 100. The buffer is filtered with Millipore filter with 0.2 μ m pore size. The NS3 protease / full length purified proteins are diluted to 4.5 nM in the assay buffer and 20 μ L of diluted protein is added to the Corning® 96 well half area black plates. The final concentration of protein in the assay was 3 nM. The different substrate concentrations ranging from 0-300 μ M were added to the enzyme and the fluorescent readings are measured at 37°C at one minute interval over five minutes using Cytation 3

Multimode plate reader (BioTek). The assays were carried out as triplicates. The AMC standard curve is determined by the relative fluorescent units against different molar concentrations of AMC. The relative fluorescent units (RFU) were converted to molar amount of AMC released according to the AMC standard curve to obtain the parameter $\text{slope}_{\text{standard}}$ (Figure 2.2B). The rate of change of fluorescent intensities was calculated by dividing the change in RFU by time interval to get the $\text{slope}_{\text{RFU}}$. Upon obtaining these two parameters, the initial velocities were calculated using the following equation (Figure 2.2B). The initial velocities are plotted using the Michaelis–Menten equation to determine V_{max} and K_m (Figure 2.2C). Rate constant (k_{cat}) was calculated from V_{max} divided by enzyme concentration $[E_t]$ (Figure 2.2D). Data were analyzed and plotted using Michaelis-Menten equation with GraphPad Prism version 5.00 for Windows, GraphPad Software, San Diego California USA.

A

B

$$\text{Initial velocity } (\mu\text{M s}^{-1}) = \text{slope}_{\text{RFU}} / \text{slope}_{\text{standard}}$$

Slope_{RFU} = rate of change of fluorescence units per second

slope_{standard} = rate of change of fluorescence unit per μM of AMC derived from AMC

standard curve

C

$$Y = V_{\max} * X / (K_m + X)$$

Y = Rates (AMC per second)

V_{max} = Maximum enzyme velocity (AMC per second per enzyme)

K_m = Michaelis–Menten constant

X = Substrate concentration (μM)

D

$$Y = E_t * k_{cat} * X / (K_m + X)$$

Y = Rates (AMC per second)

E_t = enzyme concentration

k_{cat} = rate constant

K_m = Michaelis–Menten constant

X = Substrate concentration (μM)

Figure 2.2. Scheme of protease activity assay. (A) Excitation and emission wavelengths of AMC. Figure adapted from Thermofisher Scientific. (B) Schematic diagram of fluorogenic peptide cleaved by protease releasing AMC (PeptaNova GmbH) (C) Michaelis–Menten equation used to calculate V_{max} and K_m. (D) Rate constant k_{cat} was further derived by dividing V_{max} with E_t.

Section 2.5.2: Protease Inhibition Assay

The inhibition assay was carried out using the same substrate, Bz-nKRR-AMC. The inhibitors in different concentration ranges were incubated with 3 nM protease (ZIKV

protease or DENV protease) for 30 minutes at room temperature. The reaction was initiated by addition of AMC substrate at 20 μ M. The initial velocities in RFU are recorded at 30 seconds intervals for 5 minutes. The protease with no inhibitor was used as positive control, while the protease alone with no inhibitor and substrate was used as negative control. Final assay volumes were 30 microliters. To plot the inhibition efficiency curve, the initial velocities were normalized with positive control as 100% and negative control as 0% using Normalize function in GraphPad. The inhibitor concentrations were transformed into logarithmic scale using Transform function in GraphPad Prism. Finally, the log [inhibitor] / % of inhibition function was used to determine the IC₅₀s of the inhibitors using GraphPad Prism. For high throughput screening of inhibitors, initial screening of inhibition assays were carried out with inhibitor concentrations ranging from 200-0 micromolar over 8 different inhibitor concentration. The ranges of concentrations for different inhibitors were then optimized and inhibition assays were performed over 12 different concentrations of inhibitors. Peptidomimetic inhibitors and BPTI are dissolved in assay buffer while small molecule inhibitors were dissolved in DMSO.

Section 2.5.3: Thermal shift Assays

The ThermoFluor assay was carried out as described previously (Lo et al., 2004). The assay buffer contains 20 mM HEPES pH 7.5, 150 mM NaCl, 2 mM DTT and 5% glycerol. The proteins were diluted to 10 μ M. The samples contained 10 μ M protein and 5x SYPRO Orange dye. For determination of the DENV NS3 full length protein-BPTI melting curves, 10 μ M of protein was incubated with 100 μ M of BPTI for 30 minutes before addition of the dye. The final assay volume is 50 μ L. The assay is performed in a clear 96 well PCR plate from BioRad, Bio-Rad Laboratories (Singapore) Pte. Ltd., Singapore. Using real-time PCR machine BioRad CFX96, the samples were first incubated at the temperature 25°C for 30 seconds. The samples were then heated up in temperature increments of 1°C starting from

25°C to 95°C over 20 minutes. Subsequently, the samples were cool down to 25°C. The fluorescent intensities were recorded and the baseline was corrected using the assay buffer. The melting temperatures and melting curves were calculated using Boltzmann-sigmoidal function in GraphPad Prism, version 5.00 for Windows, GraphPad Software, San Diego California USA.

Section 2.5.4: Molecular weight determination by mass spectrometry

The inhibitors were incubated with bZiPro at 10:1 ratio (inhibitor: bZiPro) for 4 hours at 37 °C. The same concentration of inhibitor without bZiPro was prepared as negative control. Analytical reverse phase chromatography was performed using ACQUITY UPLC BEH C18 column, 2.1 X 100 mm (Waters, USA), thermostated at 65 °C, with an Acquity H-Class UPLC coupled to UPLC PDA and QDa Mass Detector (Waters, USA). Mobile phase A consists of ultrapure water with 0.1 % Formic Acid. Mobile phase B consists of methanol with 0.1 % Formic Acid. The mobile phase was delivered at a flow rate of 0.4 mL/min with the following gradient: 0 – 2 min, 10 % B; 2 – 4 min, 80 % B; 4 – 6 min, 80 % B; followed by a return to 10 % B at 9 min. Absorption was measured at 280 nm and 260 nm. The QDa ESI source was operated in positive mode, capillary voltage of 0.8 V, cone voltage of 10 V, probe temperature at 600 °C, over a scan range of 150 – 1250 m/z at a frequency of 8 Hz. The chromatograms and mass spectra were acquired and analyzed using Waters Masslynx 4.1 and deconvoluted using MaxEnt1.

Section 2.5.5: ATPase assay

ATPase activity assay was carried out based on Kiianitsa et al. (Kiianitsa et al., 2003). ATP hydrolysis by the enzyme is coupled to NADH oxidation (Figure. 2.3). Depletion of NADH at 340 nm is measured using Cytation 3 Multimode plate reader (BioTek). 10 nM of enzymes were incubated in assay buffer (25 mM MOPS pH 7.4, 150 mM potassium chloride, 2 mM DTT, 0.01% Triton X-100) with 100 µM of BPTI for 30 minutes in Corning® 96 Well clear

plates. NADH mixture (NADH 1mM, Phosphoenol pyruvate 2.5mM Pyruvate Kinase 500 U/ml and Lactate dehydrogenase 100 U/ml in ATPase assay buffer) was added to reaction and incubated for 30 more minutes. Reaction is started by addition of ATP/MgCl₂ at a range of concentration from 0 – 5 mM. Absorbance at 340 nm was measured at one minute intervals for 10 minutes. Final assay volume is 50μL. Rate of ATP hydrolysis is calculated from the following equation in (Figure. 2.3). Depletion of NADH was measured by change in absorbance at 340 nm and was plotted against time. The path length for the reaction volume was calculated using Beer-Lambert law after measuring known concentration of NADH in same assay volume. After determining the path length, concentration of NADH was calculated by product of molar extinction coefficient and the path length using Beer-Lambert law. Initial velocities were calculated using the following equation from (Figure. 2.3). Rate constant k_{cat} was calculated by dividing V_{max} with enzyme concentration $[E_t]$ shown in (Figure. 2.3B). Data were plotted using Michaelis–Menten equation in GraphPad Prism version 5.00 for

Windows, GraphPad Software, San Diego California USA.

B
$$\text{Initial velocity } (\mu\text{Ms}^{-1}) = -\frac{dA_{340}}{dt} (\text{OD} \times \text{s}^{-1}) \times K^{-1}_{\text{path}}$$

K^{-1}_{path} = molar absorption coefficient for NADH at a given optical length

$$Y = V_{\max} / K_m * X$$

$$Y = \text{Rates } (\mu\text{Ms}^{-1})$$

$V_{\max}$ = maximum velocity

K_m = Michaelis-Menten constant

X = ATP concentration (μM)

$$k_{cat} = V_{\max} / [E_t]$$

Figure 2.3. ATPase activity assay. (A) Diagram showing the flow of enzymatic reactions in ATPase assay (B) ATPase activity is determined by change in absorbance for NADH divided by molar absorption coefficient of NADH at a given path length. (Kiiianitsa et al., 2003)

3. Chapter 3: Structural and biochemical characterization of ZIKV protease

Chapter 3.1: Structure determination and enzymatic characterization of ZIKV NS3 protease constructs

Section 3.1.1: Structural studies of NS3 protease

One of the recent emerging flavivirus is Zika virus (ZIKV) with the World Health Organization declaring ZIKV a world public health emergency during the 2015 outbreak in South Americas (Frieden et al., 2016; Petersen et al., 2016a). While most ZIKV infections manifest in mild symptoms such as fever, headache, rash, muscle and joint pain, ZIKV infections have been linked to microcephaly in newborns and Guillain-Barré syndrome in adults (Brasil et al., 2016; Cao-Lormeau et al., 2016; Petersen et al., 2016b). Due to its association with neurological complications in adults and infants, ZIKV poses a public health threat (Baronti et al., 2014; Bell et al., 2016; Vogel, 2016). Hence, there is still a pressing need for antiviral treatments and vaccine development against ZIKV.

Like other flaviviruses, the 11 kb (+)ssRNA genome of ZIKV encodes a polypeptide comprising three structural proteins and seven non-structural proteins (Baronti et al., 2014). The structural proteins (capsid protein C, envelope glycoprotein E, and membrane protein M) together with the lipid bilayer make up the shell of the virus particles, while the non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) are involved in the replication of the viral genome, assembly of virus particles, and evasion from host immune response. The viral RNA is first translated into a polyprotein which requires both host proteases and viral NS2B-NS3 protease, for cleavage into individual proteins. The non-structural protein NS3 is a serine protease that requires the NS2B protein as a cofactor. (Chambers et al., 1991; Chambers et al., 1990b). The hydrophilic central region of NS2B cofactor is essential for the proper folding and enzymatic activity of NS3 (Clum et al., 1997). NS2B is a transmembrane protein with its central region flanked by *N*- and *C*-terminal transmembrane regions (Chambers et al., 1993; Li et al., 2015). The NS2B transmembrane

regions are not required for enzymatic activity of NS3 but are involved in localization of NS2B-NS3 to the ER membrane (Chambers et al., 1993). The NS2B-NS3 protease cleaves the viral polyprotein at NS2A/NS2B, NS2B/NS3, NS3/NS4A, and NS4B/NS5 (Cahour et al., 1992). As the NS3 plays an indispensable role in viral replication, it is an attractive antiviral target (Kang et al., 2017).

The soluble and catalytically active recombinant construct, NS2B₄₇-G₄SG₄-NS3 protease (from here onwards to be called as **gNS2B₄₇NS3 Pro**) was designed by tethering central NS2B cofactor to NS3 protease by flexible artificial glycine linker (Leung et al., 2001). The structural studies of the NS3 protease utilized this construct design (Aleshin et al., 2007; Chandramouli et al., 2010; Erbel et al., 2006; Hammamy et al., 2013; Lei et al., 2016; Noble et al., 2012; Robin et al., 2009). These studies demonstrated that the NS2B N-terminus contributes to the folding of protease by contributing a β -strand to one of the β -barrels of protease. The C-terminus of NS2B is flexible and only observed in structures where the protease is in complex with an inhibitor or a substrate resulting in hypothesis that it is acting as a flap during substrate binding (Erbel et al., 2006; Hammamy et al., 2013; Noble et al., 2012; Robin et al., 2009). The protease apoenzyme with flexible NS2B C-terminus is said to have an “**open**” conformation while the protease-inhibitor with NS2B contributing to substrate binding site shows a “**closed**” conformation. Although all crystal structures of apo-form gNS2B₄₇NS3 protease are reported to adopt the open conformation, NMR studies have shown that gNS2B₄₇NS3 protease interchanges between open and closed conformation without the inhibitor in solution (de la Cruz et al., 2014; de la Cruz et al., 2011; Kim et al., 2013; Su et al., 2009a; Su et al., 2009b). These studies also showed that when NS2B and NS3 are separate peptides, the NS2B-NS3 protease complex is mainly in the closed conformation without the substrate (de la Cruz et al., 2014; Kim et al., 2013).

In this section, we report the structural and biochemical studies of two unlinked ZIKV NS2B₄₇NS3 protease. The two constructs are named **eZiPro**, due to the presence of NS2B/NS3 cleavage pentapeptides after NS2B C-terminus, and **bZiPro**, due to bivalent co-expression of NS2B cofactor and NS3 protease (Figure 3.1). We determine the crystal structures of eZiPro and bZiPro. We evaluate the protease activity of eZiPro and bZiPro along with conventional construct with artificial glycine linker between NS2B cofactor and NS3, named hereafter as **gZiPro**. The results here provide the structural details of ZIKV protease in post catalytic state as well as identified a new novel construct which is optimal for HTS studies and protein-inhibitor complex structure determination.

Section 3.1.2: Expression, purification and enzymatic assays of ZIKV NS3 protease constructs

We received 3 different protease constructs from Dr. Congbao Kang (ETC). The construct with the conventional glycine linker between NS2B and NS3 is named gZiPro. The one with the NS2B/NS3 cleavage pentapeptide between NS2B and NS3 is named eZiPro. The 3rd construct is named bZiPro because the NS2B cofactor and NS3 are co-expressed (Figure 3.1). The SDS-PAGE analysis of the purified proteins showed that eZiPro underwent self-proteolysis between NS2B and NS3 resulting in heterodimer similar to bZiPro (Figure 3.1B). All three proteins are monomeric in solution as observed by single peaks during the size exclusion chromatography (Figure 3.1C). The molecular weights of the proteases are confirmed by MALDI-TOF showing that there is no degradation or self-proteolysis of the protease besides the NS2B/NS3 cleavage site (Figure 3.2 A-C). The protease activities of these constructs are measured using the fluorescent substrate, Bz-nKRR-AMC (Figure 3.3A). The catalysis rates (k_{cat} s) are 1.07 ± 0.05 , 1.79 ± 0.1 and 5.30 ± 0.35 AMC molecules per second per enzyme and the apparent affinities for the substrate are 2.86 ± 0.90 , 20.42 ± 5.26 and 6.33 ± 2.41 μ M for gZiPro, eZiPro and bZiPro (Figure 3.3A). The binding of BPTI to b-,

e- and gZiPro was also measured. The half maximal inhibitory efficiencies for b-, e- and gZiPro are 12 ± 2.83 nM, 350 ± 50.90 nM and 76 ± 4.6 nM respectively. These biochemical assay results establish that bZiPro is the optimal construct for HTS screening.

Figure 3.1. Construct design and purification of ZIKV NS2B₄₇NS3 protease constructs. (A) Schematic diagram of ZIKV NS2B-NS3 gene and different ZIKV protease constructs eZiPro, gZiPro and bZiPro. The NS2B cofactor region is in magenta and hydrophobic N- and C-terminus are in grey. The NS3 is in yellow. The residues numbers at the boundaries are stated. The NS2B/NS3 cleavage pentapeptide is shown as one letter code for eZiPro. The cleavage site between NS2B and NS3 is marked with a triangle. (B) Visualization of purified gZiPro, eZiPro and bZiPro on SDS-PAGE. The gZiPro migrates to molecular weight of 25 kDa indicating the covalently linked NS2B-NS3 protease. Both eZiPro and bZiPro migrated to molecular weight between 15 and 20 kDa indicating NS3 protease alone. The reference molecular weight of the marker described next to the marker lane. (C) Size exclusion chromatography profiles of the proteins showed similar peaks indicating that large portion of the proteins are monomeric. The monomeric portion is used for structural and biochemical studies.

Figure 3.2. The molecular weights of the proteins are determined by Matrix assisted laser desorption ionization – time of flight (MALDI-TOF). The MALDI-TOF peaks for (A) gZiPro (B) eZiPro and (C) bZiPro. The expected MW and determined molecular weights are stated next to the prominent peaks.

Figure 3.3. Characterization of gZiPro, eZiPro and bZiPro enzymatic activity. (A) Protease activity of g-, e- and bZiPro are measured using Bz-nKRR-AMC substrate. The bivalent construct, bZiPro, showed highest catalytic rate while glycine-linked, gZiPro, showed lowest. The eZiPro with NS2B-NS3 cleavage site showed highest K_m . Overall, the bivalent construct shows the best enzymatic activity. (B) The half maximal inhibitory concentration of BPTI against Bz-nKRR-AMC substrate. The bZiPro has the lowest IC_{50} for BPTI. The assays were carried out as duplicates or triplicates. The error bars represent standard errors. (Phoo et al., 2016)

Section 3.1.3: Crystal structure of ZIKV NS2B-NS3 after self-cleavage (eZiPro)

We determined the crystal structure of eZiPro at 1.8 Å resolution. The refinement and data collection statistics are stated in Table 3.1. The structure consists of four molecules in two dimers in one asymmetric unit (Figure 3.4). All four molecules adopt similar fold with root mean squared deviation (RMSD) of 0.35 Å between the C-alpha atoms. The structure reveals NS2B-NS3 protease in closed conformation with NS2B cofactor wrapped around NS3 and NS2B C-terminus forming beta hairpin loop. The C-terminal residues of NS2B (from 88-95) and N-terminal residues of NS3 protease (from 1-15) are flexible and not observed in the crystal structure (Figure 3.6A). The NS2B/NS3 cleavage peptide, TGKR, is found at the

substrate pocket as indicated by the clear electron density and omit density (Figure 3.5A-D). The TGKR peptide occupy S1 and S2 pocket fully and S3 pocket partially (Figure 3.6 B, C). The carboxyl end of NS2B interacts with H51, G133 and S135 from the protease (Figure 3.6 B) The arginine side chain of R130 forms hydrogen bond with D129 and Y130. It forms cation- π interactions with Y161 as well (Figure 3.6 B). K129 lysine side chain interacts with S81 and D83 of NS2B and occupies the S2 pocket. The S3 pocket is largely unfilled due to short glycine residue at residue 128 (Figure 3.6 B). The T127 residue does not form any interaction with the protease and is located outside the pocket (Figure 3.6B). We confirmed that TGKR peptide is still covalently bound to the NS2B-cofactor by mass spectrometry (Figure 3.2C). As shown in Figure 3.2C, NS2B co-factor does not show any degradation. The residues between TGKR peptide and V87 of NS2B are flexible and are not observed in the crystal structure. After cleavage, first 18 residues of NS3 are flexible and are not observed in S' pocket.

Figure 3.4. Crystal structure of eZiPro showing (A) one unit cell with two dimers (B) one dimer of eZiPro.

Figure 3.5. NS2B C-terminus TGKR peptide in the active site of eZiPro molecules. (A-D) The TGKR peptide is shown in sticks. The electron density map $2mFo-Fc$ is in blue mesh contoured at 1σ and omit map is in green contoured at 3σ .

Figure 3.6. Crystal structure of eZiPro in complex with NS2B/NS3 cleavage peptide, TGKR. (A) Overall structure showing NS2B-NS3 protease in closed conformation with TGKR peptide at the substrate site. NS2B is colored in magenta and NS3 in yellow. The residues at the C- and N-terminus of NS2B and of NS3 are labelled with one letter code and residue number. (B) Detailed interactions between TGKR peptide and NS2B-NS3 protease. The interacting residues are labelled. The NS2B residues are underlined. Hydrogen bonds are in dashes. (C) Surface charge representation of eZiPro showing substrate binding pockets.

Section 3.1.4: Crystal structure of bivalent ZIKV protease (bZiPro)

The substrate binding pocket of eZiPro is occupied by its cleavage peptide, TGKR. Therefore, it is challenging to capture the crystal structures of ZIKV protease in complex with an inhibitor. The gZiPro free enzyme structures reported by Chen et al., showed an open NS2B conformation (Chen et al., 2016). The NMR studies of gZiPro showed that gZiPro interchanges between open and closed conformation (Li et al., 2017a). This is disadvantageous to the inhibitor binding making co-crystallization of gZiPro with inhibitors challenging. The co-expression of NS2B cofactor and NS3 has resulted in bivalent NS2B-NS3 protease, bZiPro. Unlike eZiPro, the C-terminus of NS2B cofactor does not include a self-cleavage site. Hence, the crystal structure of bZiPro would provide the optimal structure for co-crystallization with inhibitors.

Here we report the crystal structure of bZiPro at 1.58 Å resolution. The data collection and refinement statistics are shown in Table 3.1. One asymmetric unit (ASU) consists of four molecules (Figure 3.7). All four molecules are similar with RMSD of 0.45 Å between C alpha atoms. The NS2B cofactor is in closed conformation in all molecules in which the C-terminal of NS2B is folded into a beta hairpin (Figure 3.7). The four molecules are labelled as A-B, C-D, E-F and G-H (Figure 3.7). The N-terminus of G-H molecule is inserted to the substrate site of neighboring molecule A-B in adjacent ASU in a reverse orientation (Figure 3.7B). The rest of the bZiPro molecules are in closed conformation with preformed substrate binding pocket (Figure 3.7). The superimposition of bZiPro showed that it is very similar to the previous eZiPro structure (RMSD of 0.45 Å for C alpha atoms) and glycine linked NS2B-NS3 pro, gZiPro structure (RMSD of 0.52 Å for 167 C alpha atoms). This indicates that the unlinked ZIKV NS2B-NS3 protease does not undergo large conformational exchanges between free enzyme and substrate bound state (Figure 3.7).

It is interesting that the N-terminal peptide $^{14}\text{KKGE}^{17}$ binds to the neighbor bZiPro in a reverse direction. In the previous eZiPro structure, the residues 1-17 of NS3 N-terminus were disordered. In bZiPro, beginning from the T18, we can build the residues all the way to E12 unambiguously (Figure 3.7B, Figure 3.8A, B). The binding of these peptides to neighboring NS3 protease is most likely due to the crystallographic artefact since the same residues from other bZiPro molecules are flexible. The tetrapeptide $^{14}\text{KKGE}^{17}$ forms a small horseshoe like loop in the active site (Figure 3.8 A). The K14 and K15 occupy the S1 and S2 site and form most interactions with the protease (Figure 3.8A, C). The K14 act as a P1 residue and interact with D129 and Y130 residues of NS3 via hydrogen bonding (Figure 3.8A). The lysine side chain of K15 residue form hydrogen bond with S81 and D83 of NS2B while the main chain interacts with G151 and G153 of NS3 (Figure 3.8A). The G13 at P3 position does not form any interaction keeping the S3 pocket largely unoccupied similar to G129 of NS2B in eZiPro structure (Figure 3.8A, D). The last residue, E14, forms only π -stacking interaction with Y161 of NS3 protease. However, it is worthy to note that the side chain of E14 forms two intramolecular hydrogen bonds with the amino group of K14 and amide backbone of T18 (Figure 3.8 A). Similar intramolecular interactions are also observed in the structures of bZiPro with peptidomimetic inhibitors structures that are reported in Section 3.2.

Figure 3.7. Two unit cells of bZiPro. One unit cell consists of four bZiPro molecules. (A) The chains, A, C, E, and G represent NS2B and the chains, B, D, F, and H, represent NS3 protease. The N-terminus of chain H is occupying the active site of neighbor molecule in the adjacent unit cell. Of four molecules in one ASU, three have empty substrate binding sites. (B) bZiPro molecule (G:H) N-terminus is occupying the active site of the adjacent bZiPro molecule (A:B)

Figure 3.8. Crystal structure of bivalent NS2B-NS3 protease. (A) Detailed interactions between reverse peptide $^{14}\text{KKGETT}^{19}$ and active site of bZiPro. The hydrogen bonds are shown in dashes and the peptide is shown as blue sticks. The interacting bZiPro residues are labelled. The NS2B residues are underlined. (B) The electron density map of bZiPro reverse peptide $^{14}\text{KKGETT}^{19}$. The $2mF_o - F_c$ is in blue and $mF_o - F_c$ is in green contoured at 1σ and 3σ respectively. (C) Empty substrate binding pocket of bZiPro. The binding pockets, S1'-S4 are labelled. (D) The substrate binding pocket of bZiPro with reverse peptide $^{14}\text{KKGETT}^{19}$. The hydrogen bonds are shown as dashes. The $^{14}\text{KKGETT}^{18}$ peptide is shown in blue sticks. (Zhang et al., 2016a)

Table 3.1. Data collection and refinement statistics for crystal structures of ZIKV protease (eZiPro and bZiPro)

Data Collection Statistics	eZiPro	bZiPro
Wavelength (Å)	1.0004	1.0428
Resolution range (Å)	76.9- 1.84 (1.9 - 1.84)*	60.38 - 1.58 (1.60 - 1.58) *
Space group	P3 ₁	P2 ₁ 2 ₁ 2 ₁
Unit cell a, b, c, α , β , γ (Å) (°)	88.78 88.78 138.09 90.0 90.0 120.0	60.27, 60.38, 215.46, 90.0, 90.0, 90.0
Total number of reflections	499206	742298
Unique reflections	105911	108701
Multiplicity	4.7	6.8
Completeness (%)	100 (100) *	99.6 (95.4)
I/ σ I	14.1 (1.51)*	7.4 (1.8)
Wilson B-factor (Å ²)	22.2	14.7
^a R _{merge}	0.058 (0.525)*	0.134 (0.647)
Refinement Statistics		
Resolution range (Å)	31.4 - 1.84	46.17-1.58 (1.63 - 1.58)
^b R _{work} (%)	17.48	15.98
^c R _{free} (%)	20.20	19.65
Number of non-hydrogen atoms	6819	6331
Macromolecules	5883	5656
Water	936	656
Protein residues	780	761
RMSD (bonds) (Å)	0.005	0.005
^d RMSD (angles) (°)	1.19	1.13
Ramachandran favored (%)	96	96.52
Ramachandran allowed (%)	4	3.35
Ramachandran outliers (%)	0	0.13
Clashscore	12	7.24
Average B-factor (Å ²)	28.80	26.0
Macromolecules	28.60	24.9
Solvent	42.90	35.1

^a R_{merge} = $\sum I_j - \langle I \rangle / \sum I_j$, where I_j is the intensity of an individual reflection, and $\langle I \rangle$ is the average intensity of that reflection.

^b R_{work} = $\sum ||F_o| - |F_c|| / \sum |F_c|$, where F_o denotes the observed structure factor amplitude, and F_c the structure factor amplitude calculated from the model.

^c R_{free} is as for R_{work} but calculated with 5% (3044) of randomly chosen reflections omitted from the refinement.

^dRMSD = root mean square deviations

Section 3.1.5: Discussion

We report crystal structures of two different ZIKV NS2B-NS3 constructs. The first one, eZiPro, contains NS2BNS3 cleavage site pentapeptide between the NS2B cofactor sequence and NS3. The second one, bZiPro, is a bivalent construct where NS2B and NS3 are co-expressed. Both constructs give unlinked ZIKV NS2B-NS3 proteins where NS2B cofactor

and NS3 protease are two separate peptides (Figure 3.1, Figure 3.2). We characterize the enzymatic activities of these constructs and establish that bZiPro is the optimal construct for high-throughput-screening of compounds (Figure 3.3). The eZiPro structure provides post-proteolysis snapshot of ZIKV NS2B-NS3 protease. We observed that P3 residue for NS2B-NS3 cleavage site is a small amino acid, glycine which is unique to ZIKV and S3 pocket is largely unoccupied. We report the crystal structure of bivalent NS2B-NS3 construct as well. The bZiPro adopts the closed NS2B conformation. This proves that NS2B-NS3 protease is in closed conformation even before the binding of the substrate as observed by the previous NMR studies of unlinked DENV NS2B-NS3 protease (de la Cruz et al., 2014; Kim et al., 2013). This is contrary to previous flavivirus protease crystal structures reported where the NS2B cofactor is always in open conformation without a substrate (Aleshin et al., 2007; Chandramouli et al., 2010; Chen et al., 2016; Erbel et al., 2006). The open conformation observed in gZiPro construct structures could be due to the interference of artificial glycine linker between NS2B and NS3 promoting the open conformation of NS2B. The structure of the enzymatically optimal bZiPro turns out to have the empty pre-formed substrate binding site. This provides us with the opportunities for co-crystallization with different inhibitors and fragment compounds. In the next session, we will discuss our work on co-crystallization of bZiPro with different inhibitors.

Publications from this section:

Zhang, Z., Li, Y., Loh, Y. R., **Phoo, W. W.**, Hung, A. W., Kang, C., & Luo, D. (2016). **Crystal structure of unlinked NS2B-NS3 protease from Zika virus**. *Science*, 354(6319), 1597-1600. doi:10.1126/science.aai9309

Phoo, W. W. *, Li, Y. * Zhang, Z., Lee, M. Y., Loh, Y. R., Tan, Y. B., . . . Luo, D. (2016). **Structure of the NS2B-NS3 protease from Zika virus after self-cleavage**. *Nature Communications*, 7, 13410. doi:10.1038/ncomms13410

*co-first author

Author's contributions to this section:

The plasmids, gZiPro, eZiPro and bZiPro are received from Kang group, ETC, A*STAR. The protein purification, eZiPro crystal screening and eZiPro structure determination together with biochemical assays for eZiPro, bZiPro and gZiPro, were done by the author together with Ms. Zhenzhen Zhang, under supervision of Prof. Luo. The NMR studies are carried out by Kang group, ETC, A*STAR. The bZiPro crystal screening and structure determination is performed by Ms. Zhenzhen Zhang. The author contributed to the manuscript writing and making of figures.

Chapter 3.2: Exploring the substrate binding sites of the Zika protease

The development of inhibitors for flavivirus protease has been challenging due to shallow and solvent-exposed active site (Erbel et al., 2006; Hammamy et al., 2013; Noble et al., 2012). Previous crystal structures of the flavivirus NS2B-NS3 protease have been obtained with constructs containing a glycine-rich artificial linker between NS2B and NS3 (Aleshin et al., 2007; Chandramouli et al., 2010; Erbel et al., 2006; Hammamy et al., 2013; Lei et al., 2016; Noble et al., 2012; Noble and Shi, 2012; Robin et al., 2009). The structures in the absence of inhibitors showed the protease in an open conformation where the C-terminal segment of the NS2B cofactor is flexible and disordered. The linked NS2B-NS3 adopts the closed conformation only when bound with an inhibitor (Aleshin et al., 2007; Hammamy et al., 2013; Noble et al., 2012; Robin et al., 2009). However, in our previous section 3.1, we have shown that the glycine-rich artificial linker introduces steric hindrance and changes the inhibitor-binding mechanism (Figure 3.3) (Phoo et al., 2016). This conclusion has been supported by additional biochemical and NMR studies, which also showed that the unlinked construct is closer to the native state, has higher activity, and is more suitable for the identification of effective inhibitors (de la Cruz et al., 2011; Kim et al., 2013; Shannon et al., 2016). In the previous section, we have reported that bZiPro structure has an empty active site making it optimal for the co-crystallization or soaking with inhibitors.

High-throughput screening (HTS) and optimization studies have identified a number of peptidomimetic inhibitors for both DENV and WNV NS3 protease with nanomolar potency (Hammamy et al., 2013; Kouretova et al., 2017; Nitsche et al., 2012; Nitsche et al., 2013; Nitsche et al., 2017). However, for both DENV and WNV, limited availability of protein-inhibitor co-crystal structures has hindered the structure-based optimization of these inhibitors. Structural studies to complement compound identification by HTS are pivotal for the rational development of more effective antivirals. Since the structure of ZIKV NS2B-NS3

protease is very similar to the DENV and WNV proteases, previously reported inhibitors for DENV and WNV proteases could potentially be effective against the ZIKV protease. In this section, we report the crystal structures of bZiPro in complex with 2 fragment molecules, 2 dipeptide inhibitors and 3 P4-P1' peptidomimetic inhibitors along with the inhibition studies of different fragments and peptidic inhibitors.

Section 3.2.1: Crystal structures of bZiPro in complex with fragment molecules

1. EN300: Proof of concept for bZiPro as a scaffold for co-crystallization with compounds

Previously, we have demonstrated that bZiPro is the optimal construct for both high throughput biochemical assays and co-crystallization with different inhibitors since its structure contains empty preformed substrate pocket. We co-crystallized bZiPro with fragment EN300 (chemical name: 1H-benzo[d]imidazol-1-yl) methanol) and solved the structure to 2.0 Å resolution. The data collection and refinement statistics are shown in Table 3.2. The bZiPro-EN300 structure is in closed conformation and NMR experiment does not show large conformational changes upon the binding of EN300 (data not shown). The compound is partially occupying the S1 pocket by stacking with Y161 (Figure 3.9 C, D). We determined the IC_{50} of EN300 to be 14.08 mM (Figure 3.9A). It is possible that EN300 has very high IC_{50} values because it does not form any hydrogen bonds or ionic bonds with the protease and due to its small molecular weight. The thermal shift stability assays showed that upon incubation with 20 mM of EN300, T_m of bZiPro is slightly increased by 0.68 °C (Figure 3.9B).

Figure 3.9. (A) Half maximal inhibitory concentration of EN300. (B) Addition of EN300 increases the T_m of bZipro in *in vitro* thermal shift assay. (C) Overall fold of bZipro with EN300 showing the omit density. Zoomed in view of EN300 omit density map (mF_o-F_c) and electron density map ($2mF_o-F_c$) contoured at 3σ and 1σ respectively. (D) EN300 stacks with Y161. The other residues that are in vicinity of EN300 are shown as sticks and their residue numbers are labelled.

2. Pyrazole ester derivatives: Revisiting the covalent active site inhibitors of DENV

The pyrazole ester derivatives were identified during a HTS study of inhibitors against West Nile virus (WNV) protease (Johnston et al., 2007). The activity for these compounds has also been studied for DENV. Koh-Stena et al., reported three of these compounds with nanomolar and micromolar efficiency against WNV and DENV respectively (Koh-Stenta et al., 2015). Two of these compounds were found to covalently bind to the active site of the protease. We decided to explore the activity of these two small covalent inhibitors against bZipro (Figure

3.10 A,B). The chemical structures of the compounds are shown in Figure 3.10A and B. The IC_{50} of compound 1 and 5 were 1.5 μ M and 0.1 μ M respectively. As both compounds **1** and **5** are covalent inhibitors, IC_{50} values cannot accurately reflect the differences in inhibitory potential. To reveal the inhibition potency of the two compounds, we then determined the specificity constants (k_{inact}/K_i). The Compound **5** with a k_{inact}/K_i value of $0.143 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$ is significantly more potent than compound **1** which has a k_{inact}/K_i value of $0.00458 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$ (Figure 3.10 E-H). The k_{inact} of compound 5 is about 36 times higher than that of compound 1 indicating that compound 5 is able to form adduct with and inactivate the protease at a much faster rate (Figure 3.10 G,H). The covalent modification of bZiPro by compound 1 was confirmed by mass spectrometry (Figure 3.11).

Figure 3.10. (A-B) Two pyrazole ester derivative compounds, namely compound 1(A) and compound 5(B) (C,D) Half maximal inhibition efficiency (IC_{50}) of compound 1(C) and 5 (D). (E-F) Time dependent inhibition of compound 1 (E) and of compound 5 (F) show that both bind to the protease in irreversible manner. (G,H) Determination of the rate of inactivation for compound 1(G) and compound 5 (H).

Figure 3.11. Mass spectrometry analysis of bZiPro before and after incubation with compound 1. (A) Chemical structure of intact compound 1. (B) Mass increase in bZiPro after incubation with compound 1 correlated to the fragment of compound 1 shown in (C).

Crystal structure of bZiPro C143S-compound 1 complex

We report the crystal structure of bZiPro C143S in complex with compound 1 at 1.9 Å resolution. We mutated the C143S to prevent bZiPro from forming dimers. The mutation C143S does not interfere with the protease activity of bZiPro as demonstrated in Figure 3.12. There are four molecules of bZiPro in one ASU with each in complex with compound 1. The overall conformation of bZiPro-compound 1 complex shows similar fold to previous structures with RMSD 0.4 Å (Figure 3.13). The compound 1 forms a covalent bond with serine 135 (Figure 3.13A) in accordance with the mass spectrometry results. The remaining portion of the compound 1 is not found in the substrate pocket suggesting that it is released from the substrate pocket upon hydroxylation or the benzoyl moiety binds to the substrate pocket after hydrolysis. In addition the covalent bond, the compound 1 forms hydrogen bond with H51 side chain and π - π stacking interactions with Y161 and A132 (Figure 3.13A,B). It occupy S1 pocket slightly different from EN300 due to its covalent bond to the serine 135 (Figure 3.13 B, D). However, it is noteworthy that both compounds made the π - π stacking interactions with the bulky tyrosine residues surrounding the S1 site (Figure 3.9D, Figure 3.13B). These could be useful in designing more hydrophobic substrates for S1 pocket.

Figure 3.12. Protease activity of C143S and WT demonstrating that C143S mutation does not interfere with the protease activity. The error bars represent standard errors.

Figure 3.13. (A) mF_o-F_c map at 3σ and $2mF_o-F_c$ map at 1σ for compound 1 and serine 135 side chain. Compound 1 is in cyan and NS3 in yellow. (B) The compound 1 is covalently linked to S135 and forms a crucial hydrogen bond with H51. Y161 and Y150 form π - π

stacking interactions with compound 1. (C) Surface charge representation of bZiPro with compound 1 and EN300 in the S1 pocket. (D) EN300 (magenta) and compound 1(cyan) occupy the S1 pocket slightly differently. Both compounds formed π - π stacking interactions with Y161 and Y150.

Table 3.2. Data collection and refinement statistics for bZiPro-fragment molecules structures

Data Collection Statistics	bZiPro+EN300	bZiPro(C143S) + Compound1
Wavelength (Å)	1.0000	0.9537
Resolution range (Å)	43.00 - 2.00 (2.06 - 2.00)	45.74 – 1.90 (1.94 – 1.90)
Space group	P4 ₃ 22	P2 ₁ 22 ₁
Unit cell a, b, c, α, β, γ (Å) (°)	42.35, 42.35, 215.00 90.0, 90.0, 90.0	59.37, 59.63, 215.28 90.0, 90.0, 90.0
Total number of reflections	135874	367855 (22198)
Unique reflections	14167	61049 (3898)
Multiplicity	9.6	6.0 (5.7)
Completeness (%)	99.9 (99.5)	99.6 (100.0)
I/σI	14.6 (2.1)	9.9 (2.0)
Wilson B-factor (Å ²)	41.5	30.4
R _p im	0.028 (0.295)	0.058 (0.522)
CC _{1/2}	0.999 (0.925)	0.998 (0.533)
^a R _{merge}	0.071 (0.859)	0.088 (0.749)
Refinement Statistics		
Resolution range (Å)	42.34 - 2.00 (2.08 - 2.00)	42.07 – 1.90(1.81- 1.75)
^b R _{work} (%)	21.15(30.50)	23.59
^c R _{free} (%)	24.43(30.34)	27.14
Number of non-hydrogen atoms	1511	5942
Macromolecules	1430	5684
Ligands	15	32
Water	66	226
Protein residues	193	755
RMSD (bonds) (Å)	0.004	0.005
^d RMSD (angles) (°)	1.04	1.05
Ramachandran favored (%)	97.35	94.24
Ramachandran allowed (%)	2.65	5.49
Ramachandran outliers (%)	0.00	0.27
Clashscore	4.60	13.61
Average B-factor (Å ²)	55.6	29.4
Macromolecules	55.4	29.3
Ligands	74.5	24.7
Solvent	54.1	34.6

^a R_{merge} = $\sum |I_j - \langle I \rangle| / \sum I_j$, where I_j is the intensity of an individual reflection, and $\langle I \rangle$ is the average intensity of that reflection.

^b R_{work} = $\sum ||F_o| - |F_c|| / \sum |F_c|$, where F_o denotes the observed structure factor amplitude, and F_c the structure factor amplitude calculated from the model.

^c R_{free} is as for R_{work} but calculated with 5% (3044) of randomly chosen reflections omitted from the refinement.

^dRMSD = root mean square deviations

Integrity of compound 1 is necessary for binding to bZiPro

Johnston et al., has reported that compound 1 is not stable under the protease assay conditions (10 mM Tris-HCL (pH 8.0), 20% Glycerol, 0.005% Brij-35 and 1% DMSO) and hydrolyzed to 5-amino-1-[(4-methoxyphenyl) sulfonyl]pyrazol-3-ol and benzoic acid (Johnston et al., 2007). As observed from the crystal structure, the 5-amino-1-[(4-methoxyphenyl)sulfonyl]pyrazol-3-ol group is not present in the substrate pocket. To determine if benzoyl- group is necessary to bind bZiPro, compound 3 and compound 4 which are the benzoyl moiety of the compound 1, were synthesized. These compounds induce no chemical shift changes in NMR spectra indicating that integrity of compound 1 is necessary for binding to bZiPro (Figure 3.14). Therefore the degradation of the compound 1 only occurs after binding to bZiPro. Due to the presence of aromatic-aromatic interactions, Koh-Senta et al., showed that additional phenyl group to compound 1 decreased the IC_{50} (Koh-Senta et al., 2015). In our study, we showed that addition of phenyl group to the compound 1 decreased the IC_{50} by 10 times. The mass spectrometry data showed that the compound 5 binds to bZiPro irreversibly as it is the case for DENV and WNV protease (Koh-Senta et al., 2015).

Figure 3.14. ^1H - ^{15}N HSQC NMR spectra of compound 1 and its derivatives. (A) NMR spectra of free bZiPro and bZiPro + compound 1 show distinct chemical shifts indicating that compound 1 binds to bZiPro. (B) Compound 1 is degraded after binding to bZiPro to a benzoyl moiety. (C) NMR spectra of compound 3 and (D) compound 4 overlapped with bZiPro free enzyme spectra. Both compounds do not showed binding to the bZiPro indicating that just the benzoyl moiety is not enough for binding of compound 1 to bZiPro. (The NMR experiments are carried out by Kang Lab.)

Publications from this section:

Li, Y*., Zhang, Z.Z*. **Phoo, W.W***. Loh, Y.R. Li, R. Hai, Y.Y. Jansson, A.E, Hill, J. Keller, T.H. Kassoum, N. Luo, D. Kang, C.B.. (2018). **Structural insights into the inhibition of Zika virus NS2B-NS3 protease by a small molecule inhibitor**. Structure. doi: 10.1016/j.str.2018.02.005

*co-first author

Author's contributions to this section:

The inhibitors are synthesized by ETC, A*STAR. The NMR experiments are performed by Kang group at ETC, A*STAR. The mass spectrometry experiments are carried out by Kang group at ETC, A*STAR. The bZiPro C143S mutant was generated by Ms. Zhenzhen Zhang. The enzymatic activity of C143S is measured by Ms. Zhenzhen Zhang. The protein purification, co-crystallization/soaking, data collection, and structure determination of the bZiPro – EN300, bZiPro – compound 1 structures were carried out by the author together with Ms. Zhenzhen Zhang. The biochemical assays and thermal shift assays for EN300, compound 1 and compound 5 are performed by the author and Miss. Zhenzhen Zhang. The author contributed to the manuscript writing.

Section 3.2.2: Crystal structures of bZiPro in complex with dipeptides inhibitors

In this section we explore dipeptides inhibitors of with different *N-terminal* groups to characterize the S1-S2 pockets of bZiPro. The inhibitors are received from Kang Lab, ETC, A*STAR. We determined the IC₅₀s of the inhibitors (Figure 3.15). Of these inhibitors, we determined the crystal structures of bZiPro in complex with two inhibitors with nanomolar IC₅₀s.

Figure 3.15. (A-E) Evaluation of the inhibition efficiency of different dipeptide inhibitors against bZiPro using Bz-nKRR-AMC substrate. The IC₅₀ curves of the different dipeptide inhibitors are shown next to the chemical structure of the inhibitor along with the names. The error bars represent standard errors. The table shows the IC₅₀ values of the inhibitors.

Crystal structure of bZiPro with Acetyl-Lys-Arg-aldehyde (Ac-KR-H)

We report the crystal structure of bZiPro in complex with Acetyl-Lys-Arg-aldehyde (Ac-KR-H) at 2.4Å resolution. The data collection and refinement statistics are shown in Table 3.3. As expected the protease overall conformation showed the same fold with RMSD less than 0.4 Å upon superimposition with eZiPro, bZiPro and gZiPro structures (Figure 3.16A). Ac-KR-H forms a covalent bond with serine 135 and occupies the S1 and S2 pockets (Figure 3.16B,C). The arginine side chain of Ac-KR-H interacts with D129 side chain and backbone oxygen of Y130 of NS3 (Figure 3.16D). It also forms cation- π interaction with Y161 of NS3. The P2 lysine residue forms hydrogen bonds with G82 and D83 of NS2B. In addition to the covalent bond to serine, the acetyl cap interacts with G153 and H51 of NS3 (Figure 3.16D).

Figure 3.16. Crystal structure of bZiPro in complex with a dipeptide inhibitor, Acetyl-Lys-Arg-H. (A) Overall structure of bZiPro in complex with Ac-KR-H. NS2B in magenta, NS3 in yellow and Ac-KR-H in cyan. (B) Surface charge representation of bZiPro with Ac-KR-H occupying the S1 and S2 pockets. The Ac-KR-H is shown in sticks. (C) The electron density map and omit density map of Ac-KR-H and the catalytic serine 135. $mF_o - F_c$ map is contoured at 3σ in green and the $2mF_o - F_c$ map is contoured in blue at 1σ . (D) Zoom-in view of the detailed interactions of Ac-KR-H with bZiPro. The hydrogen bonds are shown in dashes, The NS2B residues in magenta and NS3 in yellow, Ac-KR-H in cyan. The residues that are interacting with Ac-KR-H are labelled. (Li et al., 2017b)

The inhibition efficiency of Ac-KR-H is about 70 times higher than the Ac-K-agmatine. To further probe the binding of these inhibitors to the bZiPro, 2D HSQC NMR spectra of bZiPro titrated with these compounds were measured (Figure 3.17). When the bZiPro is titrated with Ac-KR-H or Ac-K-Agmatine, more peaks near the active site could be observed in the NMR spectra indicating that both compounds bind to the bZiPro. The titration of Ac-KR-H to bZiPro showed that the interaction undergoes slow exchange while the titration with Ac-KR-Agmatine showed the majority of interactions undergoing fast exchange. Both spectra showed well resolved cross peaks indicating that both inhibitors are capable of stabilizing bZiPro conformation. The Ac-K-Agmatine underwent fast exchange with bZiPro due to its weak binding affinity. These results indicate that the favorable interaction between S135 and the warhead of the inhibitor should be one of the key determinants for the potency of the inhibitor.

Figure 3.17. ^1H - ^{15}N HSQC spectra of bZiPro titrated with (A) Ac-KR-H and (B) Ac-K-Argmatine. (The NMR experiments are carried out by Kang Lab.)

Crystal structure of bZiPro with imidazole-Lys-Arg-H (i-KR-H)

The crystal structure of imidazole-KR-H with bZiPro was determined to 1.7Å resolution. The data collection and refinement statistics are shown in Table 3.3. There is only one bZiPro-inhibitor complex molecule in one ASU. In this structure, the dipeptide is occupying S1 pocket partially due to its imidazole cap occupying the S1 site instead of lysine (Figure 3.18). The imidazole cap forms interactions with S135 and the backbone amide is forming hydrogen bond with G151. Due to its imidazole moiety, the i-KR-H does not form interactions with H51, D128 and Y161 as observed in bZiPro-Ac-KR-H complex (Figure 3.18D). The P1 lysine occupies the P2 pocket and forms similar interactions with bZiPro as observed in Ac-KR-H complex (Figure 3.18). The P2 Arg interacts with aspartic acid side chain of D83 and backbone amide of G153. It was unexpected that imidazole moiety would be occupying the S1 pocket; since without the crystallographic data it would be assumed that P1 Lys would be occupying the S1 pocket. When we superposed bZiPro-Acetyl-KR-H with bZiPro-imidazole-KR-H complexes, it was found that due to covalent bond formation, the peptide bond between acetyl-group and P1 Lys was rotated differently from imidazole-Lys peptide bond

(Figure 3.19). The imidazole group only forms interactions with serine 135. Interestingly, this conformation of peptide bond is observed when bZiPro is in complex with its reverse peptide KKGE (Figure 3.19). This results are further corroborated by the previous NMR experiments where stabilizing of S135 or of Y161 with fragment molecules was enough to suppress the protease exchanging from the open to closed conformation. Similar binding of the N-terminal cap moiety to the S1 pocket has been reported by Hammamy et al., in the crystal structure of WNV protease with trans-(4-guanidino)cyclohexylmethylamide cap (Hammamy et al., 2013).

Figure 3.18. Crystal structure of bZiPro in complex with a dipeptide inhibitor, imidazole-Lys-Arg-H (imidazole-KR-H). (A) Overall structure of bZiPro in complex with imidazole-KR-H. NS2B in magenta, NS3 in yellow and imidazole-KR-H in blue. (B) Surface charge representation of bZiPro with imidazole-KR-H occupying the S1 pocket partially and S2 and S3 pockets fully. The imidazole-KR-H is shown in sticks. (C) The electron density map and omit density map of imidazole-KR-H. The $mF_o - F_c$ map is contoured at 3σ in green and the $2mF_o - F_c$ map is contoured in blue at 1σ . (D) Zoom-in view of the detailed interactions of imidazole-KR-H with bZiPro. The hydrogen bonds are shown in dashes,

The NS2B residues in magenta and NS3 in yellow, imidazole-KR-H in blue.

In the new light shed by these structures, the IC_{50} s results were analyzed. We can conclude that the covalent bond between the aldehyde group and serine 135 is important for the binding of Acetyl-KR-aldehyde due to the 8 fold increase in IC_{50} when the arginine-aldehyde moiety was changed to agmatine. For imidazole-dipeptide inhibitors, we know now from the crystal structure that imidazole group is occupying the S1 pocket. Therefore, P1 Lys would be occupying S2 pocket and P2 Arg to S3 pocket. When the P2 Lys was changed to norleucine, the IC_{50} was increased by 380 folds. When P3 Arg was changed to citadine, the IC_{50} was increased by 50 folds demonstrating the importance of basic residues at P3 and P2 residues.

Figure 3.19. Comparison of the structures of bZiPro-dipeptide inhibitors with bZiPro-KKGE and with eZiPro-TGKR. (A) Superimposition of bZiPro-KKGE, eZiPro-GKR, bZiPro-Ac-KR-H and bZiPro-Imidazole-KR-H. (B) Superimposition of Ac-KR-H and imidazole-KR-H showing the different backbone conformation. (C) eZiPro-GKR and bZiPro-Ac-KR-H superimposition (D) bZiPro-KKGE reverse peptide and bZiPro-imidazole-KR-H superimposition

Table 3.3. Data collection and refinement statistics for bZiPro-dipeptide inhibitors structures

Data Collection Statistics	bZiPro+AcKR-H	bZiPro + imidazole-KR-H
Wavelength (Å)	1.0000	1.000
Resolution range (Å)	42.95 - 2.42 (2.51 - 2.42)	42.99-1.70 (1.73-1.70)
Space group	P4 ₃ 22	P4 ₁ 22
Unit cell a, b, c, α, β, γ (Å) (°)	42.33, 42.33, 214.76 90.0, 90.0, 90.0	42.61, 42.61, 214.94 90.0, 90.0, 90.0
Total number of reflections	91797 (9082)	239361 (11189)
Unique reflections	8233 (833)	22996 (1107)
Multiplicity	11.1 (10.9)	10.4 (10.1)
Completeness (%)	100.0 (99.6)	99.8 (96.3)
I/σI	12.2 (2.2)	20.6 (1.5)
Wilson B-factor (Å ²)	41.6	44.2
R _{pim}	0.026 (0.251)	0.025 (0.47)
CC _{1/2}	0.999 (0.913)	1.0 (0.829)
^a R _{merge}	0.085 (0.815)	0.059 (1.49)
Refinement Statistics		
Resolution range (Å)	42.33 - 2.42 (2.51- 2.42)	42.60 – 1.7 (1.73 – 1.70)
^b R _{work} (%)	19.12	19.72
^c R _{free} (%)	24.46	23.29
Number of non-hydrogen atoms	1474	1602
Macromolecules	1453	1476
Ligands	23	27
Water	21	99
Protein residues	193	195
RMSD (bonds) (Å)	0.008	0.016
^d RMSD (angles) (°)	0.97	1.73
Ramachandran favored (%)	95.24	96.9
Ramachandran allowed (%)	4.76	3.1
Ramachandran outliers (%)	0.00	0.00
Clashscore	4.61	4.08
Average B-factor (Å ²)	53.3	44.24
Macromolecules	53.4	43.87
Ligands	42.5	53.9
Solvent	49.0	47.14

^a $R_{\text{merge}} = \sum |I_j - \langle I \rangle| / \sum I_j$, where I_j is the intensity of an individual reflection, and $\langle I \rangle$ is the average intensity of that reflection.

^b $R_{\text{work}} = \sum ||F_o| - |F_c|| / \sum |F_c|$, where F_o denotes the observed structure factor amplitude, and F_c the structure factor amplitude calculated from the model.

^c R_{free} is as for R_{work} but calculated with 5% (3044) of randomly chosen reflections omitted from the refinement.

^d RMSD = root mean square deviations

Publication produced from this section:

Li, Y., Zhang, Z., **Phoo, W.W.**, Loh, Y.R., Wang, W., Liu, S., Chen, M.W., Hung, A.W., Keller, T.H., Luo, D. and Kang, C., (2017). **Structural dynamics of Zika virus NS2B-NS3 protease binding to dipeptide inhibitors.** Structure, 25(8), pp.1242-1250. doi:10.1016/j.str.2017.06.006

Author's contribution to this section:

The inhibitors are synthesized by ETC, A*STAR. The NMR experiments are carried out by the Kang group, ETC, A*STAR. The structure determination of bZiPro – Ac-KR-H and bZiPro – iKR-H are carried out by the author with Ms. Zhenzhen Zhang. The inhibition assays are carried out by the author together with LKC URECA student, Ms. Toh Chin Han.

Section 3.2.3: Crystal structures of bZiPro in complex with P4-P1' peptidomimetic inhibitors

Inhibition assay of peptidic inhibitors with NS3

To further characterize the ZIKV protease substrate pockets, we performed structural and biochemical studies of P4-P1' peptidic inhibitors with bZiPro. The inhibitors were received from Prof. Torsten Steinmetzer of Marburg University. The chemical structures of the inhibitors are shown in Figure 3.20. The initial IC₅₀ determination showed that MI1258 and MI1259 with cyclic backbone are most potent against bZiPro, followed by inhibitors with 4-guanidinomethyl- (4-GMe) cap at P4 and 4-amidinobenzylamide- (AMBA) cap at P1' residues.

Figure 3.20: Chemical structures of P4-P1' peptidomimetic inhibitors and inhibition assays. The inhibitors are received from Steinmetzer Lab as TFA salts. The error bars represent standard errors.

We determined the co-crystal structures of bZiPro with MI751, MI1217 and MI1229. All three molecules have shown low micromolar IC_{50} s for MI751 and MI1217 and nanomolar IC_{50} s for MI1229. During the structure refinements, the P1'-P2' residues of MI751 were not observed. To determine if bZiPro cleaves the P1' residues of the inhibitors, we performed the mass spectrometry analysis of these inhibitors (Figure 3.21). The mass spectrometry analysis showed that bZiPro cleaves off the MI751 and MI1217 between putative P1 and P1' sites. After incubation with bZiPro, the molecular weight (MW) of MI751 is decreased by 113 Daltons representing the removal of C-terminal –Gly-Gly-NH₂ segment. The MW of MI1217 is increased by 1 Dalton indicating that removal of NH₂ group and addition of OH group resulting in COOH.

Figure 3.21. Mass spectra of the compounds before and after incubation with bZiPro determined by ESI-MS. The chemical structures of the compounds before and after hydrolysis are drawn next to the peak in addition to their exact weights. (A – B) MI751 is reduced by 110 Da after incubation with bZiPro. (C – D) For MI1217, the molecular weight is increased by 1 Da after cleavage of the C-terminal amide bond. (E – F) The molecular weight of MI1229 remained unchanged before and after incubation with bZiPro.

Table 3.4. Data collection and refinement statistics for bZiPro-P4-P1 peptidomimetic inhibitors

Data Collection Statistics	bZiPro + MI751	bZiPro + MI1217	bZiPro + MI1229
Wavelength (Å)	1.0000	1.000	1.000
Resolution range (Å)	60.33 - 1.99 (2.04-1.99)	27.42 – 1.80 (1.84 -1.8)	42.99-1.70 (1.73-1.70)
Space group	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁
Unit cell a, b, c, α, β, γ (Å) (°)	60.15, 60.33, 214.07 90, 90, 90	60.33, 60.35, 214.68 90, 90, 90	60.42, 60.42, 214.87 90, 90, 90
Total number of reflections	302087 (19285)	346043 (17660)	476930 (22778)
Unique reflections	53862 (3829)	37988 (2214)	87405 (4443)
Multiplicity	5.6 (5.0)	9.1 (8.0)	5.5 (5.1)
Completeness (%)	98.40% (92.10%)	99.9% (99.60%)	99.94% (100.00%)
I/σI	10.3 (2.9)	12 (1.5)	9.3 (1.7)
Wilson B-factor (Å ²)	19.6	27.3	24.6
R _{pim}	0.026 (0.251)	0.054 (0.70)	0.025 (0.47)
CC ^{1/2}	0.99 (0.77)	0.99 (0.412)	0.99 (0.90)
^a R _{merge}	0.09 (0.44)	0.109 (1.29)	0.065(0.68)
Refinement Statistics			
Resolution range (Å)	57.907 - 1.987 (2.058 - 1.987)	24.5 - 1.67 (1.70 - 1.67)	46.181 - 1.700 (1.761 - 1.700)
^b R _{work} (%)	16.7	21	25.53
^c R _{free} (%)	20.01	26.00	29.71
Number of non-hydrogen atoms			
Macromolecules	5610	5708	5804
Ligands	78	44	116
Water	499	225	233
RMSD (bonds) (Å)	0.006	0.67	0.003
^d RMSD (angles) (°)	0.82	0.003	0.6
Ramachandran favored (%)	97.42	92.2	91.42
Ramachandran allowed (%)	0.0026	7.5	7.65
Ramachandran outliers (%)	0	0.3	0.92
Clashscore	8.12	8.03	10.63
Average B-factor (Å ²)			
Macromolecules	18.3	33.06	28.5
Ligands	22.8	34.2	31.5
Solvent	35.3	34.54	31.5

^a R_{merge} = $\sum |I_j - \langle I \rangle| / \sum I_j$, where I_j is the intensity of an individual reflection, and $\langle I \rangle$ is the average intensity of that reflection.

^b R_{work} = $\sum ||F_o| - |F_c|| / \sum |F_c|$, where F_o denotes the observed structure factor amplitude, and F_c the structure factor amplitude calculated from the model.

^c R_{free} is as for R_{work} but calculated with 5% (3044) of randomly chosen reflections omitted from the refinement.

^dRMSD = root mean square deviations

Crystal structure of bZiPro-MI751

The crystal structure of bZiPro-MI751 complex was determined at a resolution of 2 Å. The data collection and refinement statistics are summarized in Table 3.4. One ASU unit consists of four bZiPro molecules with two of them occupied by two molecules of MI751. In all four molecules, the protease is in closed conformation with NS2B C-terminus folding back into a beta hairpin loop and participating in active site. The four molecules in an ASU are essentially identical with RMSD values of 0.4 Å for 160 C-alpha atoms. The P3-P1 segment of MI751 occupies the S3-S1 pocket. The P4 phenylacetyl moiety sits out of the substrate binding pocket (Figure 3.22A, B). The P1' segment, -gly-gly-NH₂, is not observed in the crystal structure in consistent with the MS results. The oxyanion hole is intact with one of the carboxyl oxygen of P1 Arg interacting with amide nitrogen of S135 and G122. The second carboxyl oxygen interacts with imidazole of His51 (Figure 3.22B). The terminal side chain nitrogen of P1 Arg interacts with D129 side chain and carbonyl oxygen of Y130. The P2 Lys interacts with the S2 pocket via the water mediated hydrogen bonding to D75 and V72. The P3 Lys side chain comes into contact with F84 and S85 while the backbone carbonyl of P3 Lys interacts with Y161 and G153. The P4 phenyl ring is stabilized by making van der Waals interactions with V155. The horseshoe like conformation of the bound inhibitor is stabilized by two hydrogen bonds between the P4 backbone and guanidine group of P1 Arg.

Figure 3.22: Crystal structure of bZiPro in complex with MI751. (A) Surface representation of bZiPro with MI751 shown as sticks. (B) Detailed interactions of bZiPro with MI751. (C)(D) Electron density map and omit density map of two molecules of MI751. The mFo-Fc map is in green and 2mFo-Fc map is in blue contoured at 3σ and 1σ respectively. (E)(F) Intermolecular interactions of MI751. P4 backbone forms a hydrogen bond with P1 arginine side chain.

Crystal structure of bZiPro-MI1217

We also determined the co-crystal structure of bZiPro with MI1217 at a resolution of 1.8 Å. Similar to the bZiPro-compound MI751 complex, there are four bZiPro molecules in one ASU. However, two bZiPro molecules are occupied by compound MI1217 while the other two molecules have empty active sites. The RMSD between the four molecules are 0.4 Å while RMSD to previous bZiPro-MI751 complex was 0.3 Å. Again in this structure, P3-P1 residues are occupied by Lys-Lys-Arg segment residues with P4 4GMe outside the binding pocket (Figure 3.23A, B). P3-P1, Lys-Lys-Arg residues form similar interactions with NS2B-NS3 residues seen in bZiPro-MI751 complex. Notably, in the first molecule of MI1217, the

amide bond between P3 Lys and P4 4-GMe-Phac is rotated compared with the complex with compound 1, which prevents any intramolecular interaction with the P1 guanidino group (Figure 3.23 C, E). Instead, the P4 guanidine group comes into contact with the side chain of E103 of a neighboring NS3 molecule, which could cause the rotation of the amide backbone between P3 and P4 (Figure 3.23 E). The second MI1217 molecule does not interact with the neighbor NS3 molecule and forms intramolecular interactions with P1 guanidino group (Figure 3.23.D, F). The 4-GMe caps of the two molecules are slightly different as well (Figure 3.23 C and D).

Figure 3.23 Crystal structure of bZiPro in complex with MI1217. (A) Surface representation of bZiPro with MI1217 shown as sticks. (B) Detailed interactions of bZiPro with MI1217. (C)(D) Electron density map and omit density map of two molecules of MI1217. The $mF_o - F_c$ map is in green and $2mF_o - F_c$ map is in blue contoured at 3σ and 1σ respectively. (E)(F) Intermolecular interactions of MI1217. In the first molecule, the P4 guanidino group interacts with neighbor NS3 molecule. P4 backbone forms a hydrogen bond with P1 arginine side chain in second molecule. The P4 cap is in two conformations.

Crystal structure of bZiPro-MI1229

The co-crystal of bZiPro-MI1229 complex was determined at a resolution of 1.7 Å (Figure 3.13E). The data collection and refinement statistics are shown in Table 3.4. There are four molecules in one ASU, two of them in complex with MI1229. The overall crystal structure of bZiPro-MI1229 complex is similar to the already described complexes with MI751 and MI1217 with RMSD of 0.361 Å and 0.372 Å, respectively. In contrast to the two other complexes, the P1-P1' peptide bond is stable and not cleaved by bZiPro. The electron density for the P1' 4-amidinobenzylamide group is clearly visible as also supported by the results of mass spectrometry (Figure 3.21). The S3-S1 pockets are occupied by Arg-Arg-Arg segment (Figure 3.24). The P1 Arg residue binds in similar fashion as described in the complexes with MI751 and MI1217. Here, the P2 Arg side chain interacts with D83, G82 and S81 of NS2B residues. Since Arg at P2 residue forms additional interactions with the protease, it may be better suited than Lys for the design of substrate-analogue inhibitors of bZiPro (Figure 3.24). The P2 backbone adopts a slightly different conformation compared to the complexes with compounds 1 and 2, which prohibits any polar contact with bZiPro. Despite this difference, the backbone carbonyl oxygen of P3 Arg forms similar hydrogen bonds with the side chain hydroxyl of Y161 and the NH of G153 as found in the two other complexes. The P4 4-GMe-Phac moiety points out of the substrate binding pocket, whereas its guanidine group makes a

salt bridge (2.8 Å) to the carboxyl group of Asp129. Moreover, its phenyl ring partially covers the P1 side chain and is probably involved in intramolecular cation- π interactions, which stabilizes the bound inhibitor conformation. This horseshoe like overall conformation of the inhibitor is further stabilized by a hydrogen bond between the P4 carbonyl oxygen and the N ω of the P1 Arg residue. Moreover, the P1' 4-amidinobenzylamide group is not occupying the putative S1' region. Instead, it is occupying the shallow sub pocket next to the S1 site lined by V36, V52 and A132-G133, whereas the polar amine group is directed towards the solvent (Figure 3.24A and B).

Figure 3.24 Crystal structure of bZiPro in complex with MI1229. (A) Surface representation of bZiPro with MI1229 shown as sticks. (B) Detailed interactions of bZiPro with MI1229. (C)(D) Electron density map and omit density map of two molecules of MI1229. The $mF_o - F_c$ map is in green and $2mF_o - F_c$ map is in blue contoured at 3σ and 1σ respectively. (E)(F) Intermolecular interactions of MI1229. P4 backbone forms a hydrogen bond with P1 arginine side chain.

Comparison between the peptidomimetic inhibitors

The overall conformation of bZiPro peptide inhibitors showed that bZiPro does not undergo large conformational changes upon binding of inhibitors. Superimposition of bZiPro-MI series structures showed RMSD of less than 0.5 Å. The superimposition of bZiPro-MI1229 to bZiPro-TGKR peptide structure shows similar P1 interactions with horse-shoe like peptide backbone (Figure 4C). So far, five structures of flavivirus proteases in complex with peptidomimetic inhibitors containing a bulky aromatic P4 group have been determined (Erbel et al., 2006; Hammamy et al., 2013; Lei et al., 2016; Noble et al., 2012; Robin et al., 2009).

Figure 3.25. Superimposed structures of peptidic inhibitors in bZiPro, shown with its backbone in cartoon style with the NS2B colored in magenta and NS3 in yellow. (A) Superimposition of MI751 (cyan), MI1217 (orange) and MI1229 (green) in the active site of bZiPro. Conserved residues contributing to the binding of all three inhibitors are labeled. The P3-P1 residues from the three compounds make similar interactions with bZiPro, although the P2 Arg is slightly better suited than Lys at this position. The used P4 groups adopt slightly different positions indicating that the S4 pocket can be used for further optimization of the inhibitor. The P1' 4-amidinobenzylamide group of compound 3 is not occupying the S1' pocket. (B) Superimposition of the bZiPro-compound 3 complex with the structure of the boronic acid inhibitor cn-716 (PDB code 5LC0). (C) Comparison of the structure of compound 3 with the peptide TGKR after ZIKV NS2B-NS3 protease self-cleavage (PDB code 5GJ4) (D) Comparison of the structure of compound 3 and of the inhibitor 3, 4-dichlorophenylacetyl-Lys-Lys-GCMA in the active site of the WNV protease. (E) Comparison of the bZiPro-MI1229 with Bz-nKRR-H – WNV and –DENV structure (PDB id 3U1I, 2FP7) (F) Comparison of bZiPro-compound 3 to 2naphthoyl-Lys-Lys-Arg-H in complex with WNV (PDB id: 3E90)

Section 3.2.4: Discussion

In this chapter, we report several structures of ZIKV protease in complex with 1) covalently bound and non-covalently bound fragment molecules; 2) covalently and non-covalently bound di peptide inhibitors 3) non-covalently bound tetrapeptide inhibitors. We determine the potency of these compounds as well. As observed from RMSDs comparison between these structures and temperature factors, it can be concluded that bZiPro does not change conformation regardless of the size of the inhibitor. By looking at the bulky fragment molecules binding to the S1 pocket in our bZiPro-EN300 and bZiPro-compound1 structures, as well as the binding of 3,4-dichlorophenylacetyl group in S1 pocket of WNV (PDB

2YOL) (Hammamy et al., 2013), we can assume that it is possible to replace the P1 residue with bulky arginine mimics to increase the cell permeability of the drug-leads. The NMR measurements of bZiPro with compound 1 and EN300 showed that a small fragment molecule is sufficient to keep the bZiPro in the closed conformation. Subsequently, we embark on the structural characterization of the substrate binding pockets. The S3-S1 pockets are relatively conserved as the residues involved in binding P3-P1 dipeptide and Peptidomimetic inhibitors are similar for dipeptide inhibitors and tetrapeptide inhibitor (MI series). It indicates that length of peptide inhibitor does not affect the binding interactions of P1-P3. The backbone of all peptidic inhibitors adopts horseshoe like conformation in the substrate binding pocket indicating that this might be the universal backbone conformation for flaviviral protease substrates (Figure 3.19).

For longer peptidomimetic inhibitors, x-ray crystallography together with biochemical assays and mass spectrometry analysis were used to characterize the S4-S1' binding pockets of the unlinked ZIKV NS2B-NS3 protease. We analyzed the P4 groups of bZiPro-MI-series structures along with other flavivirus protease in complex with P4 capped peptidomimetic compounds (Figure 3.25). We found that P4 group has conformational freedom as seen by the different P4 moieties in different conformations and it might tolerate the incorporation of more hydrophobic groups to increase the permeability of the inhibitors. We also showed that P1' residue, AMBA in MI1229 occupy a shallow slightly hydrophobic site next to S1 pocket (Figure 3.26). This could be useful to further optimize the N-terminal capping moiety in peptidomimetic inhibitors. In structure-based drug discovery, characterizing the active site of target proteins with various compounds is vital for the further optimization of lead structures. We conclude that larger hydrophobic groups such as biphenylacetyl could be introduced to increase the cell permeability. We have identified a new and relatively hydrophobic

subpocket (Figure 3.26) which will be beneficial for the design and optimization of peptidic/peptidomimetic inhibitors for ZIKV NS2B-NS3 protease.

Figure 3.26. Surface hydrophobicity representation of ZIKV NS2B-NS3 protease. Hydrophobic and non-hydrophobic are colored red and grey respectively according to Eisenberg surface hydrophobicity scale (Eisenberg et al., 1984). Red represents the most hydrophobic regions and grey represents the least hydrophobic regions. Compound 3 is depicted in sticks and the P1', P1, P2, P3, and P4 residues labeled accordingly. A strongly hydrophobic patch is present next to S3 binding pocket.

Publication produced from the section:

Phoo, W.W.*, Zhang, Z.*, Wirawan, M.*, Chew, Edwin, Chew, B.L.A., Kouretova, J., Steinmetzer, T., Luo, D., Structures of Zika virus NS2B-NS3 protease in complex with peptidomimetic inhibitors. (Accepted in Antiviral Research)

*Co-first author

Author's contribution to this section:

The inhibitors were synthesized by Prof. Steinmetzer group from University of Marburg. The author together with Ms. Zhenzhen Zhang and Dr. Wirawan performed protein purifications, crystallization, data collection, molecular replacement, model refinement, biochemical assays under the supervision of Dr. Luo. The author together with Mr. Alvin Chew performed the mass spectrometry analysis of the inhibitors. The author contributed to the manuscript writing and figure preparations.

4. Chapter 4: Structural and biochemical characterization of NS3 from dengue virus

Chapter 4.1: Structural and biochemical characterization of NS3 protease from dengue virus

Next, we embark on our structural and biochemical studies on DENV NS3 protease. As mentioned in the previous chapter, the closed conformation of DENV protease was only captured with an inhibitor (Noble et al., 2012). With increasing evidence from NMR and biochemical studies that the unlinked NS2B cofactor NS3 protease is in closed conformation without an inhibitor, we decided to pursue the structural and biochemical studies of unlinked NS2B₄₇NS3 protease from DENV as well (Kim et al., 2013; Shannon et al., 2016). In the following section, we report the biochemical studies of three DENV NS3 protease constructs, which are similar to those of ZIKV. We name these constructs, gNS2B₄₇NS3 Pro, eNS2B₄₇NS3 Pro, and bNS2B₄₇NS3 Pro depending on the type of linker between NS2B cofactor and NS3 (Figure 4.1). We determined the crystal structure of eNS2B₄₇NS3 Pro.

Figure 4.1: Schematic drawing of NS2B-NS3 gene from DENV and construct design of NS2B₄₇NS3 Protease from DENV4.

Section 4.1.1: Construct design, cloning and purification of NS3 protease

To understand the effect of different linkers on the protease activity, we used two constructs, bNS2B₄₇NS3 Pro and eNS2B₄₇NS3 Pro similar to bZiPro and eZiPro. The eNS2B₄₇NS3 Pro was obtained by replacing the glycine linker region with five amino acids from NS2B/NS3 cleavage site by site directed mutagenesis. The bNS2B₄₇NS3 Pro was ordered from biobasic. (biobasic.sg). We also included gNS2B₄₇NS3 Pro similar to gZiPro in our studies (Figure 4.1).

The proteins were purified in the same manner for the gNS2B₄₇NS3 Pro. The SDS-PAGE analysis showed that eNS2B₄₇NS3 Pro undergoes complete proteolysis between the NS2B/NS3 junction resulting in heterodimer of NS2B cofactor and NS3 protease (Figure 4.2A and B). We determine the protease activity of different protease constructs (Figure 4.3A). The protease activity of eNS2B₄₇NS3 and bNS2B₄₇NS3 showed comparable k_{cat} s while gNS2B₄₇NS3 Pro has slightly lower k_{cat} . The eNS2B₄₇NS3 has about 2 times higher K_m than that of b- and g-NS2B₄₇NS3 Pro probably due to the presence of NS2B//NS3 cleavage junction. It can be noted that compared with ZIKV protease, the presence of NS2B/NS3 junction does not have very large reduction in the rate of catalysis (k_{cat}). The binding of BPTI to these constructs also shows only minor differences in contrast to b-, e- and gZiPro (Figure 4.3B).

Figure 4.2. (A) SDS PAGE analysis of purification of gNS2B₄₇NS3 Pro and eNS2B₄₇NS3 Pro. (B) Purified bNS2B₄₇NS3 Pro visualized on SDS-PAGE.

Figure 4.3. Characterization of enzymatic activity of the DENV protease constructs. (A) Protease activity of b-, e-, gNS2B₄₇NS3 Pro against Bz-nKRR-AMC. (B) BPTI inhibition against the protease constructs. The error bars represent standard errors.

Section 4.1.2: Crystal structure of eNS2B₄₇NS3 Pro

Figure 4.4. Pictures of eNS2B₄₇NS3 Pro crystals. (A) Crystals grown at 20°C hanging drop vapor diffusion with mother liquor 1.7 M Sodium malonate at pH 5.0. The crystals were optimized by varying the pH and dehydrated to improve the resolution. (B) Crystals grown at 20°C at sitting drop vapor diffusion with crystallization buffer 2M Sodium Malonate pH 4.0. The same dehydration method improved the crystal diffraction.

Crystal structure of eNS2B₄₇NS3 Pro

The crystals diffracted to 4 Å resolution and the resolution was slightly improved via the dehydration of the crystal overnight before freezing (Figure 4.4). The crystal structure of eNS2B₄₇NS3 Pro was determined at 3.3 Å resolution. The data collection statistics are shown in Table 4.1. The structure showed two dimers ASUs in one unit cell (Figure 4.5A). The overall fold of individual NS3 protease molecules is similar with RMSD less than 0.3 Å (Figure 4.5B). In this structure, the NS2B cofactor adopts a new conformation. So far, the reported crystal structures of NS3 protease have NS2B as either closed conformation or open conformation. In open NS2B conformation, the C-terminus of NS2B (from residues 62-89) is missing. In closed NS2B conformation, the NS2B C-terminus folds back into beta hairpin and participates in substrate binding. In our structure, the NS2B C-terminus is not fully folded back as beta hairpin (Figure 4.5B, Figure 4.6). Instead, the C-terminal NS2B/NS3 cleavage peptide residues occupy the active site of neighbor NS3 protease in adjacent unit cell. Of two dimers, the NS2B C-terminus of only one dimer (two molecules) could be modelled continuously (Figure 4.6). The two molecules belonging to the other dimer are more flexible and the electron density of NS2B was not continuous (Figure 4.6). The previous crystal structure studies of NS3 have also shown the NS2B-NS3 dimers in crystal structures (Figure 4.7). In our structure, the dimer is formed with two NS3 protease active sites facing each other in a similar manner as the gZiPro structure reported by Lei et al., (Lei et al., 2016) (Figure 4.7). Superimposition of eNS2B₄₇NS3 Pro structure on previously reported DENV protease structures showed similar overall fold of NS3 protease. The most variable regions of the protease are NS3 loop 151-158 and loop 111-121. These two loops seem to act as hinges during the cofactor movement from open to closed conformation (Figure 4.8). When we compare the surface charge density of eNS2B₄₇NS3 Pro with that of DENV2 gNS2B₄₇NS3 Pro (PDB id: 3U1I), we found that during the partial opening of NS2B

cofactor, there exists a small hydrophobic binding site between the NS2B and NS3 interface (Figure 4.9). Currently there are no constructs/ crystal structures available for capturing the protease with allosteric inhibitors. This construct/structure of NS3 protease with partially associated NS2B could be useful in co-crystallization/screening with allosteric inhibitors.

Table 4.1. Data collection and refinement statistics for eNS2B₄₇NS3 Pro

Data Collection Statistics	eNS2B ₄₇ NS3 Pro
Wavelength (Å)	
Resolution range (Å)	43.95-3.30 (3.57-3.30)
Space group	F23
Unit cell a, b, c, α, β, γ (Å) (°)	259.98 259.98 259.98 90 90 90
Total number of reflections	63319(12518)
Unique reflections	21512 (4474)
Multiplicity	2.9 (2.8)
Completeness (%)	98.1 (99.5)
I/σI	9.6 (1.5)
Wilson B-factor (Å ²)	94.5
^a R _{merge}	0.064 (0.68)
Refinement Statistics	
Resolution range (Å)	43.95 – 3.30 (3.57 – 3.30)
^b R _{work} (%)	25.60
^c R _{free} (%)	31.36
Number of non-hydrogen atoms	5858
Macromolecules	5858
Water	-
Protein residues	825
RMSD (bonds) (Å)	0.003
^d RMSD (angles) (°)	0.819
Ramachandran favored (%)	90.2
Ramachandran allowed (%)	8.8
Ramachandran outliers (%)	1
Clashscore	10.57
Average B-factor (Å ²)	
Macromolecules	130.82
Solvent	-

^a $R_{\text{merge}} = \sum |I_j - \langle I \rangle| / \sum I_j$, where I_j is the intensity of an individual reflection, and $\langle I \rangle$ is the average intensity of that reflection.

^b $R_{\text{work}} = \sum ||F_o| - |F_c|| / \sum |F_c|$, where F_o denotes the observed structure factor amplitude, and F_c the structure factor amplitude calculated from the model.

^c R_{free} is as for R_{work} but calculated with 5% (3044) of randomly chosen reflections omitted from the refinement.

^d RMSD = root mean square deviations

Figure 4.5 Crystal structure of eNS2B₄₇NS3 Pro shows novel conformation of NS2B cofactor. (A) One unit cell of eNS2B₄₇NS3 Pro showing the two dimer units. The NS3 domains are depicted in grey and NS2B in cyan, yellow, magenta and green. The chains' ids are stated next to the molecule. (B) Superimposition of four molecules of eNS2B₄₇NS3 Pro. The catalytic triads are shown as sticks.

Figure 4.6. Electron density map of NS2B co-factor for four eNS2B₄₇NS3 Pro molecules. The NS3 domains are in grey and NS2B in green, cyan, magenta and yellow. The 2mF_o-F_c

map in blue is contoured at 1σ and mF_o-F_c in green is contoured at 3σ .

Figure 4.7 NS2B-NS3 protease dimers in one ASU. (A) eNS2B₄₇NS3 protease dimer shown in cartoon (A) and in surface (B). The similar dimer was observed in gZiPro structure reported by Lei et al., (C) PDB id 5LC0. (D) In DENV3 gNS2B₄₇NS3 protease structure, the dimerization was mediated by the NS2B cofactor. PDB id 3U1I.

Figure 4.8. The NS3 loops 151-158 and 111-121 act as hinges during the NS2B-NS3 protease open and closed conformation. The NS3 protease is shown in grey, loop 151-158 and loop 111-121 are shown in yellow for eNS2B₄₇NS3 Pro, green for DENV2 gNS2B₄₇NS3 Pro closed conformation, cyan for DENV2 gNS2B₄₇NS3 Pro open conformation.

Figure 4.9. Surface charge representation of DENV protease during the partially closed conformation (A) and fully closed conformation of NS2B cofactor (B) (PDB id: 3U1I). A small hydrophobic binding site could be observed during the partial opening up stage of NS2B cofactor (A).

Section 4.1.3: Discussion

Optimizing construct designs for biochemical and structural studies are important factors in HTS and structural studies. In the chapter 3, from the bZiPro and eZiPro structures, we have established that the unlinked NS2B-NS3 protease is in closed conformation with a preformed active site. In this section we have carried out similar biochemical studies on DENV constructs, and found that the presence of NS2B/NS3 cleavage peptide between NS2B cofactor and NS3 does not affect the enzymatic activities of DENV NS3 protease as much as it does for the ZIKV NS3 protease. These findings are further proven by the similar IC_{50} values of BPTI between the three protease constructs for DENV in contrast to ZIKV. The crystal structure of eNS2B₄₇NS3 protease provides a new conformation adopted by the protease where NS2B-C-terminus is binding trans-NS3 protease complex. The structure presented here could be one of the intermediate states when the NS2B cofactor switches from

open to closed conformation. This structure could be useful in in silico modelling of the allosteric inhibitors targeting the NS2B-NS3 cofactor association sites.

Chapter 4.2: Structural and biochemical characterization of full length NS3 from dengue virus

The flavivirus non-structural protein 3 is a multi-functional protein which involves in viral poly protein processing, genome replication and viral assembly. The NS3 N-terminus (residues 1-170) is a serine protease domain and C-terminus (residue 177-618) is Helicase/NTPase domain. These two domains are connected by a flexible linker. As mentioned in chapter 3, the N-terminal protease is a serine protease which requires NS2B central cofactor region for proper folding and enzymatic activity. The C-terminal Helicase/NTPase is a DEAH box helicase SF2 family. The crystal structures of separate domains of DENV NS3 have been reported, as well as full length NS3 together with an 18-residues cofactor region of NS2B (NS2B₁₈NS3)(Chandramouli et al., 2010; Erbel et al., 2006; Luo et al., 2010; Luo et al., 2008a; Xu et al., 2005). Full length gNS2B₄₇NS3 from Murine valley encephalitis virus (MVEV) has also been reported to adopt an open conformation in the absence of inhibitor (Assenberg et al., 2009). Although the protease and NTPase/helicase domains of full length DENV4 NS2B₁₈NS3 showed similar folds to those in MVEV NS2B₄₇-NS3, domain arrangements between helicase and protease were found to be different, consistent with the flexibility of the linker region between the two activity domains (Luo et al., 2010). Here we designed bivalent, enzyme linked and conventional flexible glycine linked NS2B cofactor with full length DENV4 NS3 constructs namely bNS2B₄₇NS3, eNS2B₄₇NS3 and gNS2B₄₇NS3 similar to those in our previous studies on ZIKV protease (Phoo et al., 2016). Here we report four crystal structures of full length DENV4 NS2B₄₇NS3 constructs, eNS2B₄₇NS3 and gNS2B₄₇NS3 in apo form and two in complex with Bovine Pancreatic Trypsin Inhibitor (BPTI). The structural analysis of these NS3 constructs illustrates the dynamics of the DENV4 protease domain and its cofactor NS2B. The

biochemical studies of the enzymatic activities of full length NS3 provide the conclusive overview of uncoupled enzymatic activities of full length NS3.

Section 4.2.1: Construct design, cloning, mutagenesis and purification of full length NS3 constructs

We received the construct with 47 residues of NS2B cofactor covalently linked to NS3 full length by flexible glycine linker (to be called from here on as gNS2B₄₇NS3) from Prof. Vasudevan Lab along with two mutated constructs, gNS2B₄₇NS3 S135A and gNS2B₄₇NS3 H51N (Figure 4.10A). To overcome the problem of poor expression of wild type gNS2B₄₇NS3 proteins, we mutated the protease at either (1) S135 to alanine (S135A) or (2) hydrophobic residues on the surface, L30 and F31 to serine (L30S-F31S) (Luo et al., 2008a) (3) H51 to asparagine (H51N) (Figure 4.10A). The H51N construct has very low enzymatic activity and similar to S135A mutant and hence we did not work on this mutant any further (data not shown). The protease activity of gNS2B₄₇NS3 Pro and of gNS2B₄₇NS3 Pro (L30S-F31S) are comparable, indicating that L30S-F31S mutation does not interfere with the enzymatic activity of NS3 protease. We replaced the glycine linker of gNS2B₄₇NS3 construct with the NS2B C-terminal pentapeptide (VKTQR) resulting in endogenous enzyme cleavable NS2B-NS3 constructs - eNS2B₄₇NS3 (S135A) and eNS2B₄₇NS3 (L30S-F31S) (Figure 4.10B). The bivalent NS2B cofactor NS3 construct was designed and ordered from biobasic.sg. (Figure 4.10C)

Figure 4.10. Construct designs of DENV4 full length constructs. (A) The conventional glycine linked NS2B-NS3 full length constructs were received from Prof. Vasudevan lab. L30S F31S mutants were created by site directed mutagenesis. (B) eNS2b₄₇NS3 constructs were created by replacing the artificial glycine linker with VKTQR pentapeptide NS2B/NS3 cleavage site by SDM. (C) Bivalent bNS2B₄₇NS3 was ordered from biobasic.sg.

The protocols for recombinant protein overexpression and purification are described in chapter 2. The SDS-PAGE analysis of purified proteins is shown in Figure 2A. The eNS2B₄₇NS3 undergoes complete proteolysis between the NS2B/NS3 cleavage site resulting in separated NS2B cofactor peptide and NS3 full length, as it is the case for eNS2B₄₇NS3 Pro and eZiPro (Figure 4.11A). The SEC profiles of the purified proteins indicate them to be monomeric in solution (Figure 4.11B). To see how much internal proteolysis takes place over time, we incubated these active proteins at 37°C at 0 hr, 24 hr and 48 hr (Figure 4.7C). The

fractions were analyzed on SDS-PAGE and showed that all constructs with active protease undergoes slightly slower internal proteolysis with extra bands visible on SDS-PAGE compared to the S135A mutants (Figure 4.11C). Although the internal proteolysis happens, it is much slower than the cleavage between NS2B/NS3 junctions (Figure 4.11A). The majority of the purified NS3 stays intact (Figure 4.11C). Next, to compare the stability between different constructs, we performed *in vitro* thermal shift assay (Figure 4.11D). The T_{ms} of the free enzymes were $42.97 \pm 0.32^{\circ}\text{C}$, $43.57 \pm 0.18^{\circ}\text{C}$ and $42.48 \pm 0.25^{\circ}\text{C}$ for bNS2B₄₇NS3, eNS2B₄₇NS3 and gNS2B₄₇NS3 respectively and indicate that glycine linked construct might be slightly less stable compared to unlinked construct eNS2B₄₇NS3 and bivalent construct, bNS2B₄₇NS3 (Figure 4.11D). Upon incubation with BPTI, the T_{ms} are $44.39 \pm 0.45^{\circ}\text{C}$, $43.17 \pm 0.51^{\circ}\text{C}$ and $42.98 \pm 0.86^{\circ}\text{C}$ for bNS2B₄₇NS3, eNS2B₄₇NS3 and gNS2B₄₇NS3 respectively. The bNS2B₄₇NS3 construct has the highest increase in T_m whereas eNS2B₄₇NS3 showed no significant change in T_m . It is possible that binding of BPTI stabilizes the bNS2B₄₇NS3 while for eNS2B₄₇NS3; the presence of NS2B/3 cleavage site might cause the similar stabilization effect as the BPTI binding.

Figure 4.11. Purification of gNS2B₄₇NS3, eNS2B₄₇NS3 and bNS2B₄₇NS3 showed monomeric proteins. (A) SDS-PAGE analysis of purified full length NS3 proteins. The eNS2B₄₇NS3 undergoes complete proteolysis at NS2B/NS3 junction resulting in unlinked NS2B/NS3 full length protein similar to bNS2B₄₇NS3. (B) SEC chromatography profile of full length proteins showing that full length NS3 is monomeric. (C) SDS-PAGE analysis of full length NS3 auto proteolysis over 0 hour, 24 hour and 48 hours. (D) Melting temperatures of bNS2B₄₇NS3, eNS2B₄₇NS3 and gNS2B₄₇NS3 with/without BPTI. The error bars represent one standard error. The experiment was carried out at triplicates. The gNS2B₄₇NS3 has the lowest T_m indicating that gNS2B₄₇NS3 has lower stability compared to eNS2B₄₇NS3 and bNS2B₄₇NS3. The error bars represent standard deviations.

Section 4.2.2: Effect of L30S F31S mutation on protease activity

To determine the effect of L30 F31S mutations on the protease activity, we created gNS2B₄₇NS3 Pro WT and gNS2B₄₇NS3Pro L30S F31S by introducing STOP codon after residue 177 of NS3. Both constructs are well overexpressed in *E.coli*. The proteins are purified by using Ni-NTA affinity chromatography, followed cleavage of the His-Tag with TEV protease during the dialysis overnight. The His-Tag cleaved proteins are further purified by size exclusion chromatography. The SDS-PAGE analysis of protein purification steps showed both proteins are similar (Figure 4.12A and B). The size exclusion chromatography profiles show the proteins are monomeric (Figure 4.12C and D). We determine the protease activity of these two constructs using Bz-nKRR-AMC substrate. The protease activities are similar indicating that L30S F31S mutation does not interfere with protease activity. (Figure 4.12 E)

Figure 4.12. Overexpression and purification of DENV4 proteases, gNS2B₄₇NS3 Pro WT and its mutant L30S F31S. SDS-PAGE analysis of fractions collected during the purification steps of (A) gNS2B₄₇NS3 Pro WT and (B) gNS2B₄₇NS3 Pro L30S F31S. The size exclusion chromatography profiles of (C) gNS2B₄₇NS3 Pro WT and (D) gNS2B₄₇NS3 Pro L30S F31S. (E) Protease activity of gNS2B₄₇NS3 Pro WT and gNS2B₄₇NS3 Pro L30S F31S demonstrating that the L30S F31S mutation does not affect the catalytic activity of the proteases.

Section 4.2.3: Full length structures of NS2B₄₇NS3 from DENV4 in apo-form and in complex with BPTI

We were able to determine the crystal structures of gNS2B₄₇NS3 S135A, gNS2B₄₇NS3 L30S F31S, eNS2B₄₇NS3 S135A, and eNS2B₄₇NS3 L30S F31S in free enzyme state. We managed to capture the crystal structures of gNS2B₄₇NS3 S135A and eNS2B₄₇NS3 S135A in complex with BPTI. All four constructs crystallized in similar conditions (Table 4.2) and similar crystal morphology. The crystals were grouped according to the unit cell dimensions and conformation (Table 4.2). We observed that the unit cell dimensions of the crystal correlates with the conformation of the structure. The crystals are grouped into three different conformations, (1) Open NS2B conformation where NS2B N-terminal wraps around NS3 with NS2B C-terminus flexible, (2) Closed NS2B conformation where NS2B C-terminus folds back into β hairpin and participating in active site formation, and (3) Closed NS2B conformation in complex with BPTI. The open conformation was only captured in gNS2B₄₇NS3 (Figure 4.13 C), which closely resembles the structure of DENV4 gNS2B₁₈NS3 conformation I (PDB id: 2VBC) (Luo et al., 2008a) while the rest of the free enzyme structures are in closed conformation (Figure 4.13 C,D, F,G). The gNS2B₄₇NS3-BPTI structure and eNS2B₄₇NS3-BPTI structures adopt the same conformation (Figure 4.13E, H).

The bNS2B₄₇NS3 protein crystals on the other hand diffracted poorly to about 4Å and we failed to find a structure solution by molecular replacement. The unit cell dimensions of bNS2B₄₇NS3 crystals were similar to the closed conformations of gNS2B₄₇NS3, of linked and unlinked eNS2B₄₇NS3 which suggested that the bNS2B₄₇NS3 was also in a closed conformation (Figure 4.13 D, F, G, and Table 4.2). The data collection and refinement statistics are summarized in Table 4.3.

Construct name	Crystallisation conditions	Unit cell dimensions (a b c, α β γ)	Conformation
gNS2B₄₇NS3 S135A	0.1 M MES pH 6.4, 15% PEG 6000	52.78 75.9 86.81 89.92 90.32 93.07	Open
gNS2B₄₇NS3 L30S F31S	0.1 M MES pH 6.4, 15% PEG	52.92 88.72 81.30 90 93.08 90	Closed
eNS2B₄₇NS3 S135A	0.1M MES pH 6.4, 10% PEG 4000	52.77 88.3 81.11 90 92.34 90	Closed
eNS2B₄₇NS3 L30S F31S	0.1 M MES pH 6.4, 10% PEG 6000	52.9 88.77 81.45 90 93.85 90	Closed
gNS2B₄₇NS3 S135A + BPTI	0.1 MES pH 6.4, 12% PEG 6000	53.06 85.63 85.51 90 97.95 90	Closed, in complex with BPTI
eNS2B₄₇NS3 S135A + BPTI	0.1 M MES pH 6.0, 12% PEG 4000	53.02 87.493 86.46 90 98.25 90	Closed, in complex with BPTI
bNS2B₄₇NS3	0.1 M MES pH 6.0, 12% PEG 6000	52.65 87.64 80.12 90 91.94 90	

Table 4.2. Crystallization conditions and unit cell dimensions of NS2B₄₇NS3 constructs.

All the crystal structures of NS2B₄₇NS3 crystallized in three distinct unit cell dimensions (1) open conformation, (2) closed conformation and (3) closed in complex with BPTI. The glycine linker construct gNS2B₄₇NS3 free enzyme structures are in both open and closed conformation without an inhibitor, where eNS2B₄₇NS3 free enzyme structures are only captured in closed conformation regardless of the presence of inhibitor.

Overall, the NS2B₄₇NS3 structures adopted an extended shape where the N terminal protease and the C terminal helicase domains were connected loosely through a flexible domain linker

similar to DENV4 gNS2B₁₈NS3 structures (PDB id: 2VBC, 2WHX, 2WZQ) and the MVEV gNS2B₄₇NS3 (PDB id 2WV9) (Figure 4.13, Figure 4.14AB) (Assenberg et al., 2009; Luo et al., 2008a; Luo et al., 2008b). Compared to MVEV gNS2B₄₇NS3 structure, when helicase domains are superimposed, the protease domain rotated 177.1° degree and moved 17 Å (Figure 4.14B). In all NS2B₄₇NS3 crystals, major crystal contacts are formed between the neighboring helicase domains, which allowed the protease domain to adopt multiple conformations (Figure 4.15). From the open to the closed conformation, the protease domain translated about 0.7 Å. From the closed free enzyme conformation to the closed protease-BPTI conformation, the protease rotated 52.9° (Figure 4.13 D, E, G, H). Although of low resolution, we can still observe the solvent shells around the protease domain are discernible confirming that the structure solution is correct (Figure 4.16).

Figure 4.13. Construct design and crystal structures of DENV4 NS3. (A) Graphical representations of natural NS2B-NS3 as part of the native polyprotein, and artificial constructs discussed in this work. Construct-boundary and catalytic residues are indicated. NS2B cofactor is depicted in magenta (hydrophilic region) and grey (transmembrane regions). NS3 is represented in yellow for protease domain and cyan for helicase domain. Black arrowhead indicates site of cleavage by NS3. For eNS2B₄₇NS3 construct, five amino acid residues from NS2B/NS3 cleavage site are inserted between NS2B cofactor and NS3. For gNS2B₄₇NS3 construct, conventional artificial flexible linker (G₄SG₄) is used to covalently link the NS2B cofactor and NS3. For bNS2B₄₇NS3, each T7 promoter site is cloned in front of NS2B and NS3 resulting in co-expression of the cofactor NS2B and NS3. (B) SDS-PAGE analysis purified NS3 proteins. The first and last lanes are molecular weight markers. The construct names and mutations are indicated. The eNS2B₄₇NS3 L30S F31S and bNS2B₄₇NS3 migrated to similar size on the gel indicating the complete proteolysis between NS2B cofactor and NS3 for the eNS2B₄₇NS3 constructs. (C-H) Crystal structures of NS2B₄₇NS3 in apo-state and in complex with BPTI. The individual domains of NS3 and BPTI are labelled. The constructs used are listed below each crystal structures. Full-length NS3 adopts an elongated conformation similar to previous NS2B₁₈-NS3 full-length structure by Luo et al., (1). The NS2B cofactor is in closed conformation in both free enzyme and inhibitor bound structure for eNS2B₄₇NS3 structures (F-H).

Figure 4.14 Comparison between overall conformations of NS2B₄₇NS3 with previous structures. In both (A) and (B), gNS2B₄₇NS3 free enzyme structure in closed conformation was shown in cyan for helicase domain, yellow for protease and magenta for NS2B (A) Superposition of current full length NS2B₄₇NS3 with DENV4 full length NS3 structure with 18 residues from NS2B cofactor which is shown in blue for NS3 and magenta for NS2B (Luo et al., PDB id 2VBC) (B) Superposition of current full length NS2B₄₇NS3 structure with MVEV gNS2B₄₇NS3 structure (PDB id: 2WV9) (Assenberg et al., 2009). The MVEV NS3 is shown in orange and NS2B in magenta.

Figure 4.15. In full length NS3 structures, the major crystal contacts are formed by the helicase domain. Here we display the three conformations of gNS2B₄₇NS3 full length structures (A) Open NS2B conformation, (B) Closed NS2B conformation, (C) enzyme in complex with BPTI along with its symmetry mates. NS3 helicase domain is shown in cyan and NS3 protease in yellow. The surrounding symmetry mates are shown in grey. The NS2B is shown magenta. (D) Detailed interactions of major crystal contacts. The residues that are interacting with the symmetry mates are presented with residue number. The protease domain does not interact with neighboring molecules giving it the conformational freedom to adopt different orientations.

Figure 4.16. The $2mF_o - F_c$ maps of NS2B₄₇NS3 structures display solvent shell around the electron density of protein. This indicates that the structure solutions for the protease domain are correct and refined. The NS2B is colored in magenta, NS3 protease in yellow and BPTI in green.

Table 4.3. Data collection and refinement statistics of NS2B₄₇NS3 full length from DENV

	gNS2B ₄₇ NS3 (Open)	gNS2B ₄₇ NS3 (Closed)	eNS2B ₄₇ NS3 (linked)	eNS2B ₄₇ NS3 (unlinked)	gNS2B ₄₇ NS3 + BPTI	eNS2B ₄₇ NS3 (linked) + BPTI
Data Collection Statistics						
Wavelength (Å)	1.00	1.0004	0.9537	1.0428	1.079	1.033
Resolution range (Å)	45.05 – 2.508 (2.598 – 2.508)	45.41 - 3.097 (3.207 - 3.097)	45.27 - 3.0 (3.107 - 3.0)	42.97 - 3.2 (3.314 - 3.2)	47.69 - 2.493 (2.583 - 2.493)	45 - 2.6 (2.693 - 2.6)
Space group	P 1 2 ₁ 1	P 1 2 ₁ 1	P 1 2 ₁ 1	P 1 2 ₁ 1	P 1 2 ₁ 1	P 1 2 ₁ 1
Unit cell a, b, c, α, β, γ (Å) (°)	52.78 86.81 75.90 90 93.07 90	52.917 88.715 81.30 90 93.08 90	52.77 88.3 81.11 90 92.34 90	52.9 88.77 81.45 90 93.85 90	53.061 85.657 85.47 90 97.92 90	53.018 87.493 86.46 90 98.25 90
Total number of reflections	277395 (12526)	52184 (4835)	29692 (2981)	119885 (14348)	91122 (7831)	94028 (9737)
Unique reflections	22654 (1294)	13697 (1295)	14948 (1496)	10632 (1279)	25250 (2458)	24137 (2426)
Multiplicity	13.1 (9.7)	3.8 (3.7)	4.7 (4.8)	11.2 (11.2)	3.5 (3.2)	3.9 (4.0)
Completeness (%)	96.26 (90.82)	99.40 (96.00)	99.53 (99.93)	84.79 (99.69)	95.59 (92.97)	99.77 (99.88)
I/σI	5.5 (2.2)	13.85 (2.15)	8.55 (2.46)	6.9 (1.4)	14.03 (1.87)	11.01 (1.43)
Wilson B-factor (Å ²)	40.99	73.65	52.41	85.08	51.41	54.01
^a R _{merge}	0.097 (0.390)	0.0941 (0.5649)	0.166 (0.732)	0.078 (0.362)	0.084 (0.586)	0.0974 (0.9823)
Refinement Statistics						
^b R _{work} (%)	0.1976 (0.2791)	0.2319 (0.3528)	0.2026 (0.256)	0.2099 (0.2901)	0.2276 (0.3248)	0.2107 (0.3453)
^c R _{free} (%)	0.2458 (0.3887)	0.2782 (0.3954)	0.2480 (0.340)	0.2475 (0.3362)	0.2776 (0.3842)	0.2413 (0.3888)
Number of non-hydrogen atoms	4863	4766	4742	4686	5050	5351
Macromolecules	4733	4766	4714	4686	4940	5297
Ligand	6		15			6
Water	124		13		110	48
RMSD (bonds) (Å)	0.004	0.002	0.003	0.004	0.004	0.003
^d RMSD (angles) (°)	0.85	0.65	1.11	0.79	0.82	0.73
Ramachandran favored (%)	94	93	91	93	91	93
Ramachandran allowed (%)	5.5	6.52	8.02	6.84	8.03	6.41
Ramachandran outliers (%)	0.5	0.48	0.98	0.16	0.97	0.59
Clashscore	7.98	5.57	8.21	9.94	9.26	7.81
Average B-factor (Å ²)	53.57	82.10	61.80	101.80	74.86	75.70
NS2B	79.78	127.87	119.14	183.73	110.55	110.39

NS3	53.16	66.90	59.30	97.47	70.61	71.50
BPTI					100.82	98.78
Solvent	44.32		34.90		56.96	58.32

^a $R_{\text{merge}} = \sum |I_j - \langle I \rangle| / \sum I_j$, where I_j is the intensity of an individual reflection, and $\langle I \rangle$ is the average intensity of that reflection.

^b $R_{\text{work}} = \sum ||F_o| - |F_c|| / \sum |F_c|$, where F_o denotes the observed structure factor amplitude, and F_c the structure factor amplitude calculated from the model.

^c R_{free} is as for R_{work} but calculated with 5% (3044) of randomly chosen reflections omitted from the refinement.

^d RMSD = root mean square deviations

We captured four free enzyme full length NS2B₄₇NS3 structures. For the gNS2B₄₇NS3 construct, the NS2B cofactor is captured in both open and closed conformations (Figure 4.17 A, Figure 4.17B) while for the eNS2B₄₇NS3 constructs, the NS2B co-factor is captured only in closed conformation (Figure 4.17D, Figure 4.17E). The RMSDs of protease domain between gNS2B₄₇NS3, linked eNS2B₄₇NS3, unlinked eNS2B₄₇NS3 are less than 0.7 Å for 160 Cα atoms. In the free enzyme structures with closed NS2B conformation, the last 10-12 amino acids from the cofactor NS2B, the linker and the first 20 amino acids of NS3 are flexible. The NS2B cofactor region in the free enzyme structures shows weaker electron density compared to the full length structures in complex with BPTI (Figure 4.15). This indicates that the substrate binding stabilizes the NS2B-NS3 protease in closed conformation. Interestingly in the unlinked eNS2B₄₇NS3 structure, in contrast to the similar ZIKV protease structure, (eZiPro) (PDB id 5GJ4) (Phoo et al., 2016) reported in chapter 3, the NS2B/NS3 cleavage peptide, VKTQR is not occupying the substrate binding site.

Figure 4.17. Different conformations of NS2B cofactor in full length NS3 structures. The NS2B is shown in magenta and NS3 in yellow. The $2mF_o - F_c$ map contoured at 1σ is shown in blue and $mF_o - F_c$ map contoured at 3σ is shown in green. (A,B) The protease domain of gNS2B₄₇NS3 with cofactor NS2B shows that NS2B could be in both open and closed conformations without the inhibitor. (C,F) When the protease is in complex with BPTI, NS2B cofactor is in closed conformation for both gNS2B₄₇NS3 and eNS2B₄₇NS3 structures. (D,E) The cofactor NS2B of eNS2B₄₇NS3 protease domain stays in closed conformation without inhibitor. The electron density maps of NS2B in free enzyme structure is weaker than that of NS2B in protease-BPTI complex structure indicating that without substrate NS2B is dynamic.

Structures of full length NS3 in complex with BPTI

We also report gNS2B₄₇NS3 and linked eNS2B₄₇NS3 in complex with BPTI. The RMSD between the two NS3-BPTI structures are 0.441 Å for 618 C α atoms indicating that binding of BPTI is same in both gNS2B₄₇NS3 and eNS2B₄₇NS3 (Figure 4.10). The protease domain rotates 52.9° degree in the NS3-BPTI full length structure to accommodate the BPTI in the crystal. However, the binding did not result in major domain rearrangements. The detailed interactions between the BPTI and NS2B cofactor and NS3 protease in both structures are conserved (Figure 4.18). BPTI residues ¹²PCKAR¹⁶ and ³⁸CR³⁹ occupy the substrate binding pocket (Figure 4.18). BPTI residues K14 occupy P1 site interacting with D127, F128 while P12 and A13 of BPTI occupy P2 and P3 sites respectively. The binding of BPTI induces fully formed oxyanion hole with backbone amides of G133 and T134 interacting with carboxyl group of K14. NS2B residues ⁸⁷FS⁸⁸ interact with arginine side chain of R39 from BPTI. Electrostatic surface representation of NS2B-NS3 protease showed negatively charged P1 site (Figure 4.18B). Interestingly, residues ³⁸CR³⁹ of BPTI bind to a negative groove where R39 interacts with NS2B beta hairpin (Figure 4.18). The ATPase/helicase domains of both

gNS2B₄₇NS3 and eNS2B₄₇NS3 are similar with RMSD of less than 0.5 Å. The overall conformation of helicase is similar to the helicase structures with no NTP or RNA bound except for the residues 461-471 as mentioned before by (Assenberg et al., 2009) and residues 243-253. This surface loop is in close proximity to NS2B β hairpin, and changes conformation when the BPTI binds to the protease domain. In eNS2B₄₇NS3 -BPTI structure, movement of the protease domain results in the P-loop moving away from these residues (Figure 4.19).

Figure 4.18 Binding of BPTI to eNS2B₄₇NS3 (A) Detailed interactions of NS2B and NS3 residues and BPTI. The residues that are involved in binding are shown as sticks. The NS2B in magenta, NS3 in yellow and BPTI in green. The residues are labelled with residue numbers and one letter amino acid codes. (B) Surface representation of eNS2B₄₇NS3 in complex with BPTI binding loop shown in sticks.

Figure 4.19. The domain arrangement between protease and helicase is brought by the loop 461-471 of helicase and linker region between the helicase and the protease domain. (A) Superposition of gNS2B₁₈NS3-ADP-Mn²⁺, of free eNS2B₄₇NS3 and of eNS2B₄₇NS3 in complex with BPTI shows the conformational changes of loop 461-471 and the linker region. The helicase domains are shown in grey, NS2B₁₈NS3-ADP-Mn²⁺ protease domain is shown in blue, eNS2B₄₇NS3 is shown in green and eNS2B₄₇NS3-BPTI in red.

The artificial glycine linker interferes with the protease activity of NS3

Latest studies on biochemical and NMR studies of Zika protease have shown that the flexible glycine linker affects the enzymatic and binding activities of the protease (de la Cruz et al., 2014; Kim et al., 2013; Shannon et al., 2016). To determine the effect of artificial glycine linker on the enzymatic activities of full length NS3, we measured the protease activity of eNS2B₄₇NS3, gNS2B₄₇NS3 and bNS2B₄₇NS3 using Bz-nKRR-AMC fluorescent substrate (Li et al., 2005). Our enzymatic assays showed that while the glycine linker does not affect the substrate apparent affinity (K_m), presence of glycine linker slows down the rate of catalysis (k_{cat}) (Figure 4.20A). Although eNS2B₄₇NS3 has the NS2B/3 cleavage sequence at the NS2B-C-terminal, gNS2B₄₇NS3 has lowest k_{cat} . It is possible that the glycine linker introduces a steric hindrance on the NS2B-NS3 conformational transitions compared to the unlinked construct. Although eNS2B₄₇NS3 has a slightly higher K_m and lower k_{cat} compared to bNS2B₄₇NS3, presence of NS2B/3 cleavage site does not have the comparable inhibitory

effect as eZiPro does on the enzymatic activity of the protease as mentioned in chapter 3(Phoo et al., 2016). This could be due to the sub-optimal cleavage site found at NS2B/NS3 in all serotypes of DENV where the P2 residue is glutamine instead of a strongly basic lysine or arginine as it is for the case with other flaviviruses (Appendix 1). These results suggest that DENV NS2B-NS3 cleavage site is released from the active site upon cleavage whereas for ZIKV it remains at the active site. It is possible that different flavivirus are employing the different polyprotein cleavage site and specificity to regulate the protease activity of NS3 *in vivo*. The inhibition activity assays with BPTI and with small peptidomimetic peptidic inhibitor, Benzoyl-Lys-Arg-Arg-H, shows that the half maximal inhibitory value, IC_{50} , was lowest for bivalent bNS2B₄₇NS3 (Figure 4.20B, C) and highest for gNS2B₄₇NS3 indicating a tighter association with the former. However, these IC_{50} s are in the similar range again in contrast to ZIKV protease whereas IC_{50} of BPTI to eZiPro, gZiPro and bZiPro are 350 nM, 76 nM and 12 nM. Hence, it is plausible that in the case for DENV, the dynamics of NS2B and NS3 is involved in post-catalytic/product-release stage rather than substrate binding. This is in accordance with the single molecule enzymatic studies performed by Shannon et al., where the k_{cat} of the enzyme was affected rather than K_m when the NS2B-NS3 interactions were disrupted (Shannon et al., 2016). Having similar trends of IC_{50} s for both BPTI and Bz-NKRR-H imply that the steric hindrance caused by artificial glycine linker is independent of the size of the inhibitor. We have shown in chapter 4.1 that the Michaelis–Menten parameters and IC_{50} of the protease constructs displays similar trends to the respective full length constructs demonstrating that the effect of flexible glycine linker is independent of the helicase domain (Figure 4.3).

Figure 4.20. Characterization of protease activity of NS3 full length constructs with different linkers. (A) Protease activity of bNS2B₄₇NS3, eNS2B₄₇NS3 L30S F31S and gNS2B₄₇NS3 L30S F31S against Bz-nKRR-AMC substrate. (B,C) Half maximal inhibition efficiencies (IC_{50}) of BPTI and Benzoyl-Lys-Arg-Arg-H against NS2B₄₇NS3 constructs were determined. The gNS2B₄₇NS3 showed lowest k_{cat} and K_m . The presence of NS2B-NS3 cleavage site does not affect the enzymatic activities full length NS3 as seen by comparable k_{cat} between bNS2B₄₇NS3 and eNS2B₄₇NS3.

The kinetics of ATP hydrolysis by full length NS3 constructs are similar

Next, to determine the effect of e-, g- and bivalent linkers on the NTPase activities of NS3, we carried out the NADH coupled ATPase assay for g-, e-, bNS2B₄₇NS3 constructs. The presence of NS2B/NS3 cleavage site or glycine linker reduced the k_{cat} and increased the K_m of ATP hydrolysis slightly (Figure 4.21A). It is possible that binding of ATP could also change the relative conformation between protease and helicase as observed by Luo et al (Luo et al., 2008b). The presence of glycine linker might be restricting these movements and hence increasing the K_m whereas the presence of NS2B/NS3 cleavage site at C-terminus of NS2B might be interfering with ATP hydrolysis by coming into contact with substrate pocket. To test the effect of BPTI binding to ATPase activity, we measured the ATPase activity of bNS2B₄₇NS3 in the presence and absence of BPTI. The DENV4 Helicase domain alone was used as the control. The rate of ATP hydrolysis remains unchanged when BPTI is bound to the protease domain demonstrating that the substrate binding on protease domain does not affect the ATPase activity of helicase domain (Figure 4.21B). The ATPase activity of DENV4 NS3 helicase was measured in the presence and absence of BPTI as the control (Figure 4.21B). Compared to helicase, both k_{cat} and K_m of ATP hydrolysis of bNS2B₄₇NS3 is slightly slower compared to those of helicase alone, and hence the catalytic efficiencies of both enzymes are similar (Figure 4.21B).

Figure 4.21. The ATP hydrolysis activities of NS2B₄₇NS3 constructs were compared along with Helicase. (A) Rate of ATP hydrolysis of bNS2B₄₇NS3, eNS2B₄₇NS3 and gNS2B₄₇NS3. The Michaelis–Menten parameters are stated below the curves. The presence of artificial glycine linker and of NS2BNS3 cleavage junction slightly lowers the ATP hydrolysis by increasing K_m . (B) ATP hydrolysis of bNS2B₄₇NS3 with or without BPTI shows that presence of inhibitor does not interfere with ATP hydrolysis. The DENV4 helicase was used as positive control.

Section 4.2.4: Discussion

Due to lack of crystal structures in closed conformation without substrate or inhibitor, NS2B was proposed to change from open and closed conformations depending on substrate binding. Early NMR studies of glycine linked DENV and WNV proteases showed crowded cross peaks due to conformational exchanges (de la Cruz et al., 2011; Su et al., 2009a). The use of unlinked constructs has improved the spectral quality and backbone assignment (de la Cruz et al., 2014; Kim et al., 2013). The unlinked DENV protease constructs are obtained by 1) replacement of glycine linker with NS2B/NS3 cleavage site (EVKKQR) similar to eNS2B₄₇NS3 and 2) by co-expression of NS2B cofactor peptide and NS3 protease similar to

bNS2B₄₇NS3 (de la Cruz et al., 2014; Kim et al., 2013; Shannon et al., 2016). The NMR studies of these unlinked DENV proteases have confirmed that the NS2B cofactor is predominantly in closed conformation without the substrate. The NMR studies for similar ZIKV protease, eZiPro, bZiPro and gZiPro, also showed that unlinked protease is in closed conformation (Li et al., 2017a; Phoo et al., 2016; Zhang et al., 2016b). These studies also proved that the higher population of open conformation of NS2B₄₇NS3 is mainly due to the glycine linker (de la Cruz et al., 2014; Kim et al., 2013; Li et al., 2017a). We have established for ZIKV the NS2B-NS3 protease is in closed conformation with or without substrate. We have also established that the presence of the artificial linker introduces a steric hindrance in the protease conformational changes. Based on these results, in this section, we test if these findings are also applicable to the full length NS3. We characterize the full length NS3 from DENV4 with different linker constructs. In general, these full-length constructs formed monomers in solution as indicated by the size exclusion chromatography. The overall structures confirm that the protease domain is mainly in closed conformation in full length context as well. Combined with biochemical assays, we conclude that the artificial glycine linker present between NS2B cofactor and NS3 could be interfering with the protease activity by promoting the conformational heterogeneity of the protease. In contrast to ZIKV protease, the binding of inhibitors does not make a difference for different DENV full length constructs. For both linked and unlinked eNS2B₄₇NS3, the NS2B/NS3 cleavage site pentapeptide (VKTQR) is not found at the active site, different from similar protease structure from ZIKV, eZiPro discussed in chapter 3.1. We also characterize the ATPase activities of full length NS3. The ATPase activities of full length NS3 and of helicase domain alone are similar. Binding of BPTI does not interfere with the ATPase activity of helicase domain further indicating that the enzymatic activities of protease and helicase domains are uncoupled. All the structures reported here are crystallized under similar crystallization

conditions (Table 4.2). Because helicase domain is responsible for major crystal contacts, the protease domain has relatively free conformations (Figure 4.14). It is possible to use the helicase domain as the scaffold for the crystallization of different proteases from other flavivirus such as JEV, or YFV.

From our structures, we propose that NS2B-NS3 protease mainly stays in closed conformation with or without substrate. During the polyprotein processing, the NS2B is anchored to ER membrane by N and C-terminal hydrophobic regions (Clum et al., 1997). The complete dissociation of NS2B C-terminus from NS3 protease would not be favorable spatially due to the NS2B membrane anchorage whereas stable tight association of whole NS2B cofactor to NS3 will put the active site of NS3 closely to the membrane, shielding it from the substrate binding or interfering with substrate release. Therefore, loosely associated NS2B-cofactor would be the optimal conformation for NS2B-NS3 *in vivo*.

In conclusion, we provide the structural evidence that the NS2B cofactor loosely assumes closed conformation with NS3 protease in full length NS3 without substrate. Unlike eZiPro, the substrate pocket of eNS2B₄₇NS3 is not occupied and therefore may be useful for co-crystallization with inhibitors. Due to better protease activities, bNS2B₄₇NS3 and eNS2B₄₇NS3 are better suited for high-throughput screening of potential drugs.

Author contributions and publications produced from this chapter

Phoo, W.W., Sahili, A.E., Zhang, Z., Chen, M.W., Liew, C.W., Lescar, J.*, Vasudevan, S.V.*, Luo, D.* (2018). Crystal structures of full length NS3 from Dengue virus provide insights into dynamics of protease domain. (Manuscript in preparation)

*corresponding authors

The author performed protein crystallization, construct mutagenesis, data collection and refinement, biochemical assays under the supervision of Dr. Luo. The author contributed to the manuscript writing and figure preparations.

5. Conclusions and future directions

Conclusions

- From this study, we determined the crystal structures of two ZIKV protease constructs. The empty pocket of bZiPro structure opened new opportunities for soaking and co-crystallization with different inhibitors. Seizing this opportunity we subsequently determined the crystal structures of bZiPro in complex with two small molecule inhibitors, one covalently bound and another non-covalently bound; with two dipeptide inhibitors and with peptidomimetic inhibitors with different N- and C-termini. The structural information provided by these structures will be used for optimization of inhibitors in structure based drug discovery.
- We determined the crystal structure of DENV eNS2B₄₇NS3 Pro. The structure exhibits a partially open NS2B cofactor conformation of the protease which has never been captured before. We also determined three DENV NS3 full length free enzyme structures and two DENV NS3 full length structures in complex with BPTI.
- The biochemical activities of b-, e-, gZiPro and of b-, e-, g-NS2B₄₇NS3 Pro showed that the presence of artificial linker interferes with the protease activity. The comparison between the protease activities of b-, e-, g-ZiPro and of b-, e-, gNS2B₄₇NS3 Pro demonstrated that either NS2B/NS3 cleavage sites from different species of flaviviruses behave differently, or that substrate binding and product release of the NS3 protease from different flaviviruses behave uniquely.
- We determined four full length NS3 structures from DENV4 in free enzyme states and two full length NS3 structures in complex with BPTI. These structures established that the protease domain is mainly in closed conformation regardless of the inhibitor binding in full length context as well. The NS3-BPTI structures showed that the binding of BPTI to the protease domain of full length NS3 is conserved. The different linkers between NS2B cofactor and NS3 protease have effect on the protease activity but not on the ATPase activity in the full length NS3. The protease and ATPase activities of NS3 are uncoupled and binding of protease to substrate does not interfere with the ATPase activity of the helicase domain.

Future directions

The future directions for this study could be separated into two paths, towards the discovery of antivirals against the flaviviruses and towards the understanding the roles of NS3 during the viral replication.

- We have determined the co-crystal structures of bZiPro in complex with in total seven inhibitors. Yet, there are many non-peptidic inhibitors that were identified by a number of studies as reviewed in Chapter 1. Capturing the bZiPro with these bigger non-peptidic inhibitors will be of great value to the structure based drug discovery. The structural information of how these non-peptidic inhibitors bind to the active site could give us how we could depeptidize our current peptidomimetic inhibitor. The first generation of HCV protease inhibitors were substrate based/peptidomimetic inhibitors that were depeptidized step by step. The same strategy could be applied for either specific or pan-flavi protease inhibitors as well. It is notable that HCV protease inhibitor drug discovery utilized about 20 HCV NS3 protease inhibitor complex structures.
- Another approach for the antiviral against NS3 is the allosteric inhibitors. Studies by Brecher et al., and Li et al., have shown potent allosteric inhibitors of ZIKV protease in enzymatic and cell-based assays as well as in mouse model(Brecher et al., 2015; Li et al., 2017c). A bZiPro construct with shorten NS2B N-terminus could be designed to capture the bZiPro in complex with these allosteric inhibitors. In case of the bZiPro NS2B cofactor N-terminus interferes with crystallization, we can also use the DENV4 NS3 full length construct with shorten NS2B-N terminus too. Modification of the linker between NS2B and NS3 does not interfere with the crystallization of the protein as proven by our DENV4 NS3 full length free enzyme structures. In fact, in our eNS2B47NS3 S135A structure, the first 5 residues from N-terminus of NS2B were not associated with NS3 and flexible. This construct might be useful for structural determination of DENV NS3 full length with the allosteric inhibitors.
- During the viral replication *in vivo*, the NS3 is anchored to the membrane by NS2B N- and C-terminus. So far only NMR studies of membrane associated NS2B is available but overall conformation of NS3 in complex with membrane anchor NS2B is missing. It will be interesting the study the biochemical and structural studies of full length / membrane associated NS2B-NS3 protease/full length.

6. Appendix 1: Multiple sequence alignment of NS2B and NS3 from flavivirus

Multiple sequence alignment of non-structural protein 3 from flavivirus

DENV4 $\beta 22$ $\alpha 7$ $\beta 23$ $\alpha 8$ $\alpha 9$ $\beta 24$ $\eta 2$
 360 370 380 390 400 410
 DENV4 GKT VWF VPS I K A G N D I A N C L R K S G K K V I Q L S R K T F D T E Y P K T K L T D W D F V V T T D I S E M G A
 DENV3 GKT VWF VPS I K A G N D I A N C L R K N G K K V I Q L S R K T F D T E Y Q K T K L N D W D F V V T T D I S E M G A
 DENV2 GKT VWF VPS I K T G N D I A A C L R K N G K R V I Q L S R K T F D S E Y V K T R T N D W D F V V T T D I S E M G A
 DENV1 GKT VWF VPS I K S G N D I A N C L R K N G K R V V Q L S R K T F D T E Y Q K T K N N D W D Y V V T T D I S E M G A
 JEV GKT VWF V A S V K M G N E I A M C L O R A G K K V I Q L N R K S Y D T E Y P K C K N G D W D F V I T T D I S E M G A
 WNV GKT VWF VPS V K M G N E I A L C L O C A G K K V V Q L N R K S Y E T E Y P K C K N D W D F V I T T D I S E M G A
 YFV R P T A W F L P S I R A A N V M A A S L R K A G K S V V V L N R K T F E R E Y P T I K Q K K P D F I L A T D I S E M G A
 ZIKV GKT VWF VPS V R N G N E I A A C L T K A G K R V I Q L S R K T F E T E F Q K T K H Q E W D F V V T T D I S E M G A
 MVEV GKT VWF V A S V K M S N E I A Q C L O R A G K R V I Q L N R K S Y D T E Y P K C K N G D W D F V I T T D I S E M G A

DENV4 $\beta 25$ $\beta 26$ $\beta 27$ $\alpha 10$ $\beta 28$
 420 430 440 450 460 470
 DENV4 N F R A G R V I D E R R C L K P V I L P D G P E R V I L A G P I P V T P A S A A Q R R G R I G R N P A Q E D D Q Y V F S
 DENV3 N F K A D R V I D E R R C L K P V I L T D G P E R V I L A G P M P V T A S A A Q R R G R V G R N P Q K E N D Q Y I F T
 DENV2 N F K A E R V I D E R R C M K P V I L T D G E E R V I L A G P M P V T H S A A A Q R R G R I G R N P R N E N D Q Y I Y M
 DENV1 N F R A D R V I D E R R C L K P V I L K D G P E R V I L A G P M P V T V A S A A Q R R G R I G R N Q N K E G D Q Y I Y M
 JEV N F G A S R V I D C R K S V K P T I L E E G E G R V I L G N P S P I T S A S A A Q R R G R V G R N P N Q V G D B Y H Y G
 WNV N F K A S R V I D S R K S V K P T I I T E G E G R V I L G E P S A V T A S A A Q R R G R I G R N P S Q V G D B Y C Y G
 YFV N L C E R V I D G R T A F K K P V L V D E G R . K V A I K G P L R I S A S A A Q R R G R I G R N P N R D G D S Y Y S
 ZIKV N F K A D R V I D S R R C L K P V I L D . . G E R V I L A G P M P V T H A S A A Q R R G R I G R N P N K P G D B Y L Y G
 MVEV N E G A S R V I D C R K S V K P T I L D E G E G R V I L S V P S A I T S A S A A Q R R G R V G R N P S Q I G D B Y H Y G

DENV4 $\alpha 11$ $\alpha 12$ $\alpha 13$
 480 490 500 510 520 530
 DENV4 G D P L K N D E D H A H W T E A K M L D N I Y T P E G I I P T L F G P E R E K T Q A I D G E F R L R G E Q R K T F V E
 DENV3 G Q P L N N D E D H A H W T E A K M L D N I N T P E G I I P A L F E P E R E K S A A I D G E Y R L K G E S R K T F V E
 DENV2 G E P L N D E D C A H W K E A K M L D N I N T P E G I I P S I F E P E R E K V D A I D G E Y R L R G E A R K T F V D
 DENV1 G Q P L N N D E D H A H W T E A K M L D N I N T P E G I I P A L F E P E R E K S A A I D G E Y R L R G E A R K T F V E
 JEV G A T S E D D S N L A H W T E A K I M L D N I H M P N G L V A Q L Y G P E R E K A F T M D G E Y R L R G E E K K N F L E
 WNV G H T N E D D S N F A H W T E A R I M L D N I N M P N G L I A Q F Y Q P E R E K V Y T M D G E Y R L R G E E R K N F L E
 YFV E P T S E N N A H H V C L E A S M L D N M E V R G G M V A P L Y G V E G T K T P V S P G E M R L R D D Q R K V F R E
 ZIKV G G C A E T D E D H A H W L E A R M L D N I Y L Q D G L I A S L Y R P E A D K V A A I B G E F K L R T E Q R K T F V E
 MVEV G T S E D D T M L A H W T E A K I L D N I H L P N G L V A Q L Y G P E R D K T Y T M D G E Y R L R G E E R K T F L E

DENV4 $\eta 3$ $\eta 4$ $\beta 29$ $\beta 30$
 540 550 560 570 580 590
 DENV4 L M R R G D L P V W L S Y K V A S A G I S Y E D R E W C F T G E R N N Q I L E E N M E V . E I W T R E G E K K L R P R
 DENV3 L M R R G D L P V W L A H K V A S E G I K Y T D R K W C F D G O R N N Q I L E E N M D V . E I W T K E G E K K L R P R
 DENV2 L M R R G D L P V W L A Y K V A A E G I N Y A D R R W C F D G T R N N Q I L E E N V E V . E I W T K E G E R K K L R P R
 DENV1 L M R R G D L P V W L S Y K V A S E G F Q Y S D R R W C F D G E R N N Q V L E E N M D V . E I W T K E G E R K K L R P R
 JEV L L R T A D L P V W L A Y K V A S N G I Q Y T D R R W C F D G P R T N A I L E D N T E V . E I V T R M G E R K I L R P R
 WNV L L R T A D L P V W L A Y K V A A A G V S Y H D R R W C F D G P R T N T I L E D N N E V . E V I T K L G E R K T L R P R
 YFV L V R N C D L P V W L S W Q V A K A G L K T N D R K W C F E G P E E H E I L N D S G E T V K C R A P G G A K K L R P R
 ZIKV L M K R G D L P V W L A Y Q V A S A G I T Y T D R R W C F D G T N N T I M E D S V P A . E V W T R H G E K R V L K P R
 MVEV L I K T A D L P V W L A Y K V A S N G I Q Y N D R K W C F D G P R S N I I L E D N N E V . E I I T R I G E R K V L K P R

DENV4 $\eta 5$ $\alpha 15$
 600 610
 DENV4 W L D A R V Y A D P M A L K D F K E F A S G R K
 DENV3 W L D A R T Y S D P L A L K E F K D F A A G R K
 DENV2 W L D A R I Y S D P L A L K E F K E F A A G R K
 DENV1 W L D A R T Y S D P L A L R E F K E F A A G R R
 JEV W L D A R V Y A D H Q A L K W F K D F A A G K R
 WNV W I D A R V Y S D H Q A L K A F K D F A S G K R
 YFV W C D E R V S D Q S A L S E F I K F A B G R R
 ZIKV W M D A R V C S D H A A L K S F K E F A A G K R
 MVEV W L D A R V S D H Q S L K W F K D F A A G K R

Multiple sequence alignment of non-structural protein 2B from flavivirus

DENV4

1 10 20 30 40 50 $\beta 1$

DENV4 .WP LNEGIMAVGLVSLLGSA LLKN DVP . LAGPMVAGGLLLAAYVMSGSSADLSLEKAANV
DENV3 ... LNEGVMAVGLVSLILASS LLRN DVP . MAGPLVAGGLLIACYVITGTSADLTVEKAANV
DENV2 ... LNEAIMAVGMVSLILASS LLKN DIP . MTGPLVAGGLLTVCYVLTGRSADLELERAADV
DENV1 ... LNEGIMAIGIVSILLSS LLKN DVP . LAGPLIAGGMLIACYVISGSSADLSLEKAAEV
JEV . . PATEFLSAVGLMFAIVGG LAELDIESMSIPFMLAGLMAVSYVVS GKATDMWLERAADI
WNV . . PATEFLSAVGLMFAIVGG LAELDIESMSIPFMLAGLMAVSYVVS GKATDMWLERAADI
YFV SIPVNEALAAAGLVGLAG . LAFQEMENFLGPIAVGGLLMMLVSVAGRVDGLELKKLGEV
ZIKV SWPPSEVLTAVGLICALAGGFAKADIE . MAGPMAAVGLLIVSYVVS GKSVDMYIERAGDI
MVEV GWPATEVLTAVGLMFAIVGG LAELDIDMSVPTIAGLMLVSYVIS GKATDMWLERAADV

DENV4

60 70 80 90 100 110 $\beta 2$ $\beta 3$

DENV4 QWDEMAADITGSSPIIEVKQDEDFGSFSIRDV EETNMITL LVKLALITVSGLYPLAIPVTMA
DENV3 TWEDEAEQTVSHNLMITVD DDGTMR IKD DETENILTV LLKTALLIVSGIFPYSIPATLL
DENV2 KWEDQAEISGSSPILSITISEDGSMSEIKNEEEQTLTI LIRTGLLVISGVFPVSIPIITAA
DENV1 SWEDEAEHSGASHNILLVEVQDDGTMKIKDEERDDTLTI LLKATLLAISGVYPMSSIPATLF
JEV SWEMDAAITGSSRRRLDVKLD DDGDFHLIDDPGV PWKVV VLRMSCIGLAALTPWAI VPAAF
WNV SWEMDAAITGSSRRRLDVKLD DDGDFHLIDDPGV PWKVV VLRMSCIGLAALTPWAI VPAAF
YFV SWEDEAEISGSSARYDVALSEQGEFKLLSEKVPWDQV VMTSLALVGAALHPFALLLVLA
ZIKV TWEKDAEVTGNSPRLDVALDESGDFS LVEDDGP PMREI ILKVVLMTICGMNP IATPFAAG
MVEV SWEAGAAITGTSERLDVQLDDGDFHL LNDP GV PWKI W VLRMTCLSVAAITPRAITLPSAF

DENV4

120

DENV4 L WYIWQVKTQR
DENV3 V HHTWQKQTQR
DENV2 A WYLWEVKKQR
DENV1 V WYFWQKKKQR
JEV G YWLTTLKTTKR
WNV G YWLTTLKTTKR
YFV G WLFHVRGARR
ZIKV A WYVYVKTGKR
MVEV G YWLTTLKYTKR

7. Appendix Table 1. Amino acid sequences of the constructs used in crystallography

Table 7.1. Amino acid sequences of the constructs used in crystallography

ZIKV NS2B-NS3 protease constructs	
bZiPro	MGSSHHHHHHSSGLVPRGSGALWDVPAPKEVKKGETTDGVYRVMTRRLLGSTQVG VGVMQEGVFHTMWHVTKGAALRSGEGRDPYWGDKQLVSYCGPWKLDAAW DGLSEVQLLAVPPGERAKNIQTLPGIFKTKDGDIGAVDYPAGTSGSPILDKCGRVI GLYGNGVVIKNGSYVSAITQGKREEETPVE* MTGKSVDMYIERAGDITWEKDAEVTGNSPRLDVALDESGDFSLVEEDGPPMRE*
eZiPro	GSSHHHHHHSSGLVPRGSHMTGKSVDMYIERAGDITWEKDAEVTGNSPRLDVALDE SGDFSLVEEDGPPMREKTGKRSALWDVPAPKEVKKGETTDGVYRVMTRRLLGST QVGVGVMQEGVFHTMWHVTKGAALRSGEGRDPYWGDKQLVSYCGPWKLDA AWDGLSEVQLLAVPPGERAKNIQTLPGIFKTKDGDIGAVDYPAGTSGSPILDKCG RVIGLYGNGVVIKNGSYVSAITQGKREEETPVE*
gZiPro	MGSSHHHHHHSSGLVPRGSHMTGKSVDMYIERAGDITWEKDAEVTGNSPRLDVAL DESGDFSLVEEDGPPMREGGGSGGGSGALWDVPAPKEVKKGETTDGVYRVMTR RLLGSTQVGVGVMQEGVFHTMWHVTKGAALRSGEGRDPYWGDKQLVSYCGP WKLDAAWDGLSEVQLLAVPPGERAKNIQTLPGIFKTKDGDIGAVDYPAGTSGSPI LDKCGRVIGLYGNGVVIKNGSYVSAITQGKREEETPVE*
DENV NS2B-NS3 protease constructs	
bNS2b ₄₇ NS3 Pro	MHHHHHHSSGVDLGTEENLYFQSMGSGALWDVPSAATQKAALSEGVYRIMQRGLF GKTQVGVIHIEGVFHTMWHVTRGSVICHETGRLEPSWADVRNDMISYGGGWRLG DKWDKEEDVQVLAIEPGKNPKHVQTKPGLFKTLTGEIGAVTLDFKPGTSGSPIINRK GKVIGLYGNGVVTKSGDYVSAITQAERIGEPDYE* MGGSADLSLEKAANVQWDEMADITGSSPIIEVKQDEDEGSFSIRDVEETNMI*
eNS2B ₄₇ NS3 Pro	MHHHHHHSSGVDLGTEENLYFQGAMGSADLSLEKAANVQWDEMADITGSSPIIEVK QDEDEGSFSIRDIEETNMIAQVKTKRSGALWDVPSAATQKATLSEGVYRIMQRGLFG KTQVGVIHIMAGVFHTMWHVTRGSVICHETGRLEPSWADVRNDMISYGGGWRLG DKWDKEEDVQVLAIEPGKNPKHVQTKPGLFKTLTGEIGAVTLDFKPGTSGSPIINKK GKVIGLYGNGVVTKSGDYVSAITQAERIGEPDYEVED*
gNS2B ₄₇ NS3 Pro	MHHHHHHSSGVDLGTEENLYFQGAMGSADLSLEKAANVQWDEMADITGSSPIIEVK QDEDEGSFSIRDVEETNMIGGGSGGGSGALWDVPSAATKKAALSEGVYRIMQRG LFGKTQVGVIHMEGVFHTMWHVTRGSVICHETGRLEPSWADVRNDMISYGGGWRL LGDKWDKEEDVQVLAIEPGKNPKHVQTKPGLFKTLTGEIGAVTLDFKPGTSGSPIINR KGKVIGLYGNGVVTKSGDYVSAITQAERIGEPDYEVED*
DENV NS2B-NS3 full length constructs	
bNS2b ₄₇ NS3	MGGSSADLSLEKAANVQWDEMADITGSSPIIEVKQDEDEGSFSIRDVEETNMIT* MHHHHHHSSGVDLGTEENLYFQSMGSGALWDVPSAATKKAALSEGVYRIMQRGLF

GKTQVGVGIIHMEGVFHTMWHVTRGSVICHETGRLEPSWADVRNDMISYGGGWRL
 GDKWDKEEDVQVLAIEPGKNPKHVQTKPGLFKTLTGEIGAVTLDFKPGTSGSPIINR
 KGKVIGLYGNGVVTKSGDYVSAITQAERIGEPDYEVEDEDIFRKKRLTIMDLHPGAGK
 TKRILPSIVREALKRRLRTLILAPTRVVAAEMEEALRGLPIRYQTPAVKSEHTGREIVD
 LMCHATFTTRLLSSTRVPNYNLIVMDEAHFTDPSSVAARGYISTRVEMGEAAAFMT
 ATPPGATDPFPQSNPIEDIEREIPERSWNTGFDWITDYQGKTVWFVPSIKAGNDIANC
 LRKSGKKVIQLSRKTFDTEYPKTKLTDWDFVVTDDISEMGANFRAGRVIDPRRCLKP
 VILPDGPERVILAGPIPVTPASAAQRRGRIGRNPAQEDDQYVFSGDPLKNDEDHAHW
 TEAKMLLDNIYTPEGIPTLFGPEREKTQAIDGFEFRLRGEQRKTFVELMRRGDLPVWL
 SYKVASAGISYEDREWCFGTGERNNQILEENMEVEIWTREGEKKKLRPRWLDARVYA
 DPMALKDFKEFASGRK*

eNS2B₄₇NS3 MSYYHHHHHDYDIPTTENLYFQGAMGSADLSLEKAANVQWDEMADITGSSPIIEV
 KQDEDGSFSIRDIEETNMIVKTQRSGALWDVPSPAATQKATLSEGVYRIMQRGLFGK
 TQVGVGIIHMEGVFHTMWHVTRGSVICHETGRLEPSWADVRNDMISYGGGWRLGD
 KWDKEEDVQVLAIEPGKNPKHVQTKPGLFKTLTGEIGAVTLDFKPGTSGSPIINKKG
 KVIGLYGNGVVTKSGDYVSAITQAERIGEPDYEVEDEDIFRKKRLTIMDLHPGAGKTK
 RILPSIVREALKRRLRTLILAPTRVVAAEMEEALRGLPIRYQTPAVKSDHTGREIVDL
 MCHATFTTRLLSSTRVPNYNLIVMDEAHFTDPCSVAARGYISTRVEMGEAAAFMTA
 TPPGSIDPFPQSNPIEDIEREIPERSWNTGFDWITDYQGKTVWFVPSIKAGNDIANCL
 RKSGKRVIQLSRKTFDTEYPKTKLTDWDFVVTDDISEMGANFRAGRVIDPRRCLKPV
 ILTDGPERVILAGPIPVTPASAAQRRGRIGRNPAQEDDQYVFSGDPLKNDEDHAHW
 EAKMLLDNIYTPEGIPTLFGPEREKTQAIDGFEFRLRGEQRKTFVELMRRGDLPVWL
 SYKVASAGISYEDREWCFGTGERNNQILEENMEVEIWTREGEKKKLRPRWLDARVYA
 DPMALKDFKEFASGRK*

gNS2B₄₇NS3 MSYYHHHHHDYDIPTTENLYFQGAMGSADLSLEKAANVQWDEMADITGSSPIIEV
 KQDEDGSFSIRDIEETNMIGGGGSGGGGSGALWDVPSPAATQKATLSEGVYRIMQR
 GLFGKTQVGVGIIHMEGVFHTMWHVTRGSVICHETGRLEPSWADVRNDMISYGGGW
 RLGDKWDKEEDVQVLAIEPGKNPKHVQTKPGLFKTLTGEIGAVTLDFKPGTSGSPIIN
 KKGKVIGLYGNGVVTKSGDYVSAITQAERIGEPDYEVEDEDIFRKKRLTIMDLHPGAG
 KTKRILPSIVREALKRRLRTLILAPTRVVAAEMEEALRGLPIRYQTPAVKSDHTGREIV
 DLMCHATFTTRLLSSTRVPNYNLIVMDEAHFTDPCSVAARGYISTRVEMGEAAAFMT
 TATPPGSIDPFPQSNPIEDIEREIPERSWNTGFDWITDYQGKTVWFVPSIKAGNDIAN
 CLRKSGKRVIQLSRKTFDTEYPKTKLTDWDFVVTDDISEMGANFRAGRVIDPRRCLK
 PVILTDGPERVILAGPIPVTPASAAQRRGRIGRNPAQEDDQYVFSGDPLKNDEDHAH
 WTEAKMLLDNIYTPEGIPTLFGPEREKTQAIDGFEFRLRGEQRKTFVELMRRGDLPV
 WLSYKVASAGISYEDREWCFGTGERNNQILEENMEVEIWTREGEKKKLRPRWLDARV
 YADPMALKDFKEFASGRK*

8. Appendix Table 2: List of crystal structures reported in the thesis

Table 8.1. List of NS2B-NS3 structures reported in this thesis.

Construct name	Ligand	PDB code
ZIKV		
eZiPro	TGKR	5GJ4
bZiPro	-/KKGE	5GPI
bZiPro	EN300	5H4I
bZiPro	Ac-KR-H	5H6V
bZiPro	i-KR-H	NA
bZiPro	Ac-K-agmatine	NA
bZiPro	MI751	5ZMQ
bZiPro	MI1217	5ZMS
bZiPro	MI1229	5ZOB
DENV		
eNS2B ₄₇ NS3 Pro	-	NA
eNS2B ₄₇ NS3 L30S F31S		5YVW
eNS2B ₄₇ NS3 S135A		5YVU
gNS2B ₄₇ NS3 S135A		5YVJ
gNS2B ₄₇ NS3 L30S F31S		5YVV
eNS2B ₄₇ NS3 S135A	BPTI	5YW1
gNS2B ₄₇ NS3 L30SF31S	BPTI	5YVY

9. References

- . Baltimore Classification of Viruses. In Molecular Biology Web Book.
- Abbink, P., Larocca, R.A., De La Barrera, R.A., Bricault, C.A., Moseley, E.T., Boyd, M., Kirilova, M., Li, Z.F., Ng'ang'a, D., Nanayakkara, O., *et al.* (2016). Protective efficacy of multiple vaccine platforms against Zika virus challenge in rhesus monkeys. *Science* 353, 1129-1132.
- Abola, E., Kuhn, P., Earnest, T., and Stevens, R.C. (2000). Automation of X-ray crystallography. *Nature Structural Biology* 7, 973-977.
- 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. *J Biol Chem* 276, 39926-39937.
- Acosta, E.G., Castilla, V., and Damonte, E.B. (2009). Alternative infectious entry pathways for dengue virus serotypes into mammalian cells. *Cell Microbiol* 11, 1533-1549.
- Adachi, O., Kawai, T., Takeda, K., Matsumoto, M., Tsutsui, H., Sakagami, M., Nakanishi, K., and Akira, S. (1998). Targeted disruption of the MyD88 gene results in loss of IL-1- and IL-18-mediated function. *Immunity* 9, 143-150.
- Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66, 213-221.
- Afonine, P.V., Grosse-Kunstleve, R.W., Chen, V.B., Headd, J.J., Moriarty, N.W., Richardson, J.S., Richardson, D.C., Urzhumtsev, A., Zwart, P.H., and Adams, P.D. (2010a). phenix.model_vs_data: a high-level tool for the calculation of crystallographic model and data statistics. *J Appl Crystallogr* 43, 669-676.
- Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., and Adams, P.D. (2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* 68, 352-367.
- Afonine, P.V., Mustyakimov, M., Grosse-Kunstleve, R.W., Moriarty, N.W., Langan, P., and Adams, P.D. (2010b). Joint X-ray and neutron refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* 66, 1153-1163.
- Aguiar, M., Halstead, S.B., and Stollenwerk, N. (2017). Consider stopping dengvaxia administration without immunological screening. *Expert Review of Vaccines* 16, 301-302.
- Aguiar, M., Stollenwerk, N., and Halstead, S.B. (2016). The risks behind Dengvaxia recommendation. *Lancet Infectious Diseases* 16, 882-883.
- Akey, D.L., Brown, W.C., Dutta, S., Konwerski, J., Jose, J., Jurkiw, T.J., DelProposto, J., Ogata, C.M., Skiniotis, G., Kuhn, R.J., *et al.* (2014). Flavivirus NS1 structures reveal surfaces for associations with membranes and the immune system. *Science* 343, 881-885.
- Aleshin, A.E., Shiryayev, S.A., Strongin, A.Y., and Liddington, R.C. (2007). Structural evidence for regulation and specificity of flaviviral proteases and evolution of the Flaviviridae fold. *Protein Sci* 16, 795-806.
- Alexopoulou, L., Holt, A.C., Medzhitov, R., and Flavell, R.A. (2001). Recognition of double-stranded RNA and activation of NF-kappaB by Toll-like receptor 3. *Nature* 413, 732-738.
- 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. *J Virol* 68, 3794-3802.
- Anglero-Rodriguez, Y.I., Pantoja, P., and Sariol, C.A. (2014). Dengue virus subverts the interferon induction pathway via NS2B/3 protease-IkappaB kinase epsilon interaction. *Clinical and vaccine immunology : CVI* 21, 29-38.
- Apte-Sengupta, S., Sirohi, D., and Kuhn, R.J. (2014). Coupling of replication and assembly in flaviviruses. *Curr Opin Virol* 9, 134-142.
- Arjona, A., Ledizet, M., Anthony, K., Bonafe, N., Modis, Y., Town, T., and Fikrig, E. (2007). West Nile virus envelope protein inhibits dsRNA-induced innate immune responses. *J Immuno* 179, 8403-8409.

Arroyo, J., Miller, C., Catalan, J., Myers, G.A., Ratterree, M.S., Trent, D.W., and Monath, T.P. (2004). ChimeriVax-West Nile virus live-attenuated vaccine: Preclinical evaluation of safety, immunogenicity, and efficacy. *J Virol* 78, 12497-12507.

Artenstein, A.W. (2012). The discovery of viruses: advancing science and medicine by challenging dogma. *Intl J Infect Dis* 16, E470-E473.

Ashour, J., Laurent-Rolle, M., Shi, P.Y., and Garcia-Sastre, A. (2009). NS5 of dengue virus mediates STAT2 binding and degradation. *J Virol* 83, 5408-5418.

Ashraf, U., Ye, J., Ruan, X., Wan, S., Zhu, B., and Cao, S. (2015). Usutu virus: an emerging flavivirus in Europe. *Viruses* 7, 219-238.

Assenberg, R., Mastrangelo, E., Walter, T.S., Verma, A., Milani, M., Owens, R.J., Stuart, D.I., Grimes, J.M., and Mancini, E.J. (2009). Crystal structure of a novel conformational state of the flavivirus NS3 protein: implications for polyprotein processing and viral replication. *J Virol* 83, 12895-12906.

Avirutnan, P., Hauhart, R.E., Somnuk, P., Blom, A.M., Diamond, M.S., and Atkinson, J.P. (2011). Binding of flavivirus nonstructural protein NS1 to C4b binding protein modulates complement activation. *J Immuno* 187, 424-433.

Avirutnan, P., Punyadee, N., Noisakran, S., Komoltri, C., Thiemmecca, S., Auethavornanan, K., Jairungsri, A., Kanlaya, R., Tangthawornchaikul, N., Puttikhunt, C., *et al.* (2006). Vascular leakage in severe dengue virus infections: a potential role for the nonstructural viral protein NS1 and complement. *The J Infect Dis* 193, 1078-1088.

Ayres, C., Guedes, D., Paiva, M., Donato, M., Barbosa, P., Krokovsky, L., Rocha, S., Saraiva, K., Crespo, M., Rezende, T., *et al.* (2017). Response to: 'Lack of evidence for Zika virus transmission by Culex mosquitoes'. *Emerg Microbes Infect* 6, e91.

Bardina, S.V., Bunduc, P., Tripathi, S., Duehr, J., Frere, J.J., Brown, J.A., Nachbagauer, R., Foster, G.A., Krysztof, D., Tortorella, D., *et al.* (2017). Enhancement of Zika virus pathogenesis by preexisting anti-flavivirus immunity. *Science* 356, 175-180.

Barkhash, A.V., Perelygin, A.A., Babenko, V.N., Brinton, M.A., and Voevoda, M.I. (2012). Single nucleotide polymorphism in the promoter region of the CD209 gene is associated with human predisposition to severe forms of tick-borne encephalitis. *Antiviral research* 93, 64-68.

Baronti, C., Piorkowski, G., Charrel, R.N., Boubis, L., Leparac-Goffart, I., and de Lamballerie, X. (2014). Complete coding sequence of Zika virus from a French polynesia outbreak in 2013. *Genome Announc* 2.

Baronti, C., Sire, J., de Lamballerie, X., and Querat, G. (2010). Nonstructural NS1 proteins of several mosquito-borne Flavivirus do not inhibit TLR3 signaling. *Virology* 404, 319-330.

Battye, T.G., Kontogiannis, L., Johnson, O., Powell, H.R., and Leslie, A.G. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* 67, 271-281.

Beckett, C.G., Tjaden, J., Burgess, T., Danko, J.R., Tamminga, C., Simmons, M., Wu, S.J., Sun, P.F., Kochel, T., Raviprakash, K., *et al.* (2011). Evaluation of a prototype dengue-1 DNA vaccine in a Phase 1 clinical trial. *Vaccine* 29, 960-968.

Behnam, M.A., Graf, D., Bartenschlager, R., Zlotos, D.P., and Klein, C.D. (2015). Discovery of Nanomolar Dengue and West Nile virus Protease Inhibitors Containing a 4-Benzoyloxyphenylglycine Residue. *J Med Chem* 58, 9354-9370.

Behnam, M.A., Nitsche, C., Vechi, S.M., and Klein, C.D. (2014). C-terminal residue optimization and fragment merging: discovery of a potent Peptide-hybrid inhibitor of dengue protease. *ACS Med Chem Lett* 5, 1037-1042.

Bell, B.P., Boyle, C.A., and Petersen, L.R. (2016). Preventing Zika Virus Infections in Pregnant Women: An Urgent Public Health Priority. *Am J Public Health* 106, 589-590.

Benarroch, D., Selisko, B., Locatelli, G.A., Maga, G., Romette, J.L., and Canard, B. (2004). The RNA helicase, nucleotide 5'-triphosphatase, and RNA 5'-triphosphatase activities of Dengue virus protein NS3 are Mg²⁺-dependent and require a functional Walker B motif in the helicase catalytic core. *Virology* 328, 208-218.

Bera, A.K., Kuhn, R.J., and Smith, J.L. (2007). Functional characterization of cis and trans activity of the Flavivirus NS2B-NS3 protease. *J Biol Chem* 282, 12883-12892.

Bessaud, M., Pastorino, B.A., Peyrefitte, C.N., Rolland, D., Grandadam, M., and Tolou, H.J. (2006). Functional characterization of the NS2B/NS3 protease complex from seven viruses belonging to different groups inside the genus Flavivirus. *Virus Res* 120, 79-90.

Best, S.M., Morris, K.L., Shannon, J.G., Robertson, S.J., Mitzel, D.N., Park, G.S., Boer, E., Wolfinbarger, J.B., and Bloom, M.E. (2005a). Inhibition of interferon-stimulated JAK-STAT signaling by a tick-borne flavivirus and identification of NS5 as an interferon antagonist. *J Virol* 79, 12828-12839.

Best, S.M., Morris, K.L., Shannon, J.G., Robertson, S.J., Mitzel, D.N., Park, G.S., Boer, E., Wolfinbarger, J.B., and Bloom, M.E. (2005b). Inhibition of interferon-stimulated JAK-STAT signaling by a tick-borne flavivirus and identification of NS5 as an interferon antagonist. *J Virol* 79, 12828-12839.

Blacksell, S.D. (2012). Commercial Dengue Rapid Diagnostic Tests for Point-of-Care Application: Recent Evaluations and Future Needs? *J Biochem Biomed*

Blitvich, B.J., Scanlon, D., Shiell, B.J., Mackenzie, J.S., Pham, K., and Hall, R.A. (2001). Determination of the intramolecular disulfide bond arrangement and biochemical identification of the glycosylation sites of the nonstructural protein NS1 of Murray Valley encephalitis virus. *J Gen Virol* 82, 2251-2256.

Bollati, M., Alvarez, K., Assenberg, R., Baronti, C., Canard, B., Cook, S., Coutard, B., Decroly, E., de Lamballerie, X., Gould, E.A., *et al.* (2010). Structure and functionality in flavivirus NS-proteins: perspectives for drug design. *Antiviral Res* 87, 125-148.

Brasil, P., Pereira, J.P., Jr., Moreira, M.E., Ribeiro Nogueira, R.M., Damasceno, L., Wakimoto, M., Rabello, R.S., Valderramos, S.G., Halai, U.A., Salles, T.S., *et al.* (2016). Zika Virus Infection in Pregnant Women in Rio de Janeiro. *N Engl J Med* 375, 2321-2334.

Brecher, M., Chen, H., Liu, B., Banavali, N.K., Jones, S.A., Zhang, J., Li, Z., Kramer, L.D., and Li, H. (2015). Novel Broad Spectrum Inhibitors Targeting the Flavivirus Methyltransferase. *PloS one* 10, e0130062.

Brinkworth, R.I., Fairlie, D.P., Leung, D., and Young, P.R. (1999). Homology model of the dengue 2 virus NS3 protease: putative interactions with both substrate and NS2B cofactor. *The J Gen Virol* 80 (Pt 5), 1167-1177.

Brinton, M.A. (1986). Replication of Flaviviruses. In *The Togaviridae and Flaviviridae*, S. Schlesinger, and M.J. Schlesinger, eds. (Boston, MA: Springer New York), pp. 327-374.

Brownlie, J. (1990). Pathogenesis of Mucosal Disease and Molecular Aspects of Bovine Virus Diarrhea Virus. *Vet Microbiol* 23, 371-382.

Bunkoczi, G., Echols, N., McCoy, A.J., Oeffner, R.D., Adams, P.D., and Read, R.J. (2013). Phaser.MRage: automated molecular replacement. *Acta Crystallogr D Biol Crystallogr* 69, 2276-2286.

Byrd, C.M., Grosenbach, D.W., Berhanu, A., Dai, D., Jones, K.F., Cardwell, K.B., Schneider, C., Yang, G., Tyavanagimatt, S., Harver, C., *et al.* (2013). Novel benzoxazole inhibitor of dengue virus replication that targets the NS3 helicase. *Antimicrob Agents Chemother* 57, 1902-1912.

Cahour, A., Falgout, B., and Lai, C.J. (1992). Cleavage of the dengue virus polyprotein at the NS3/NS4A and NS4B/NS5 junctions is mediated by viral protease NS2B-NS3, whereas NS4A/NS4B may be processed by a cellular protease. *J Virol* 66, 1535-1542.

Campbell, G.L., Hills, S.L., Fischer, M., Jacobson, J.A., Hoke, C.H., Hombach, J.M., Marfin, A.A., Solomon, T., Tsai, T.F., Tsu, V.D., *et al.* (2011). Estimated global incidence of Japanese encephalitis: a systematic review. *Bulletin of the World Health Organization* 89, 766-774.

Canduri, F., and de Azevedo, W.F. (2008). Protein Crystallography in Drug Discovery. *Current Drug Targets* 9, 1048-1053.

Cao-Lormeau, V.M., Blake, A., Mons, S., Lastere, S., Roche, C., Vanhomwegen, J., Dub, T., Baudouin, L., Teissier, A., Larre, P., *et al.* (2016). Guillain-Barre Syndrome outbreak associated with Zika virus infection in French Polynesia: a case-control study. *Lancet* 387, 1531-1539.

Cao, X., Li, Y., Jin, X., Li, Y., Guo, F., and Jin, T. (2016). Molecular mechanism of divalent-metal-induced activation of NS3 helicase and insights into Zika virus inhibitor design. *Nucleic Acids Res* 44, 10505-10514.

Capeding, M.R., Tran, N.H., Hadinegoro, S.R.S., Ismail, H.I.H.J.M., Chotpitayasunondh, T., Chua, M.N., Luong, C.Q., Rusmil, K., Wirawan, D.N., Nallusamy, R., *et al.* (2014). Clinical efficacy and safety of a novel tetravalent dengue vaccine in healthy children in Asia: a phase 3, randomised, observer-masked, placebo-controlled trial. *Lancet* **384**, 1358-1365.

Chambers, T.J., Grakoui, A., and Rice, C.M. (1991). Processing of the yellow fever virus nonstructural polyprotein: a catalytically active NS3 proteinase domain and NS2B are required for cleavages at dibasic sites. *J Virol* **65**, 6042-6050.

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

Chambers, T.J., Nestorowicz, A., Amberg, S.M., and Rice, C.M. (1993). Mutagenesis of the yellow fever virus NS2B protein: effects on proteolytic processing, NS2B-NS3 complex formation, and viral replication. *J Virol* **67**, 6797-6807.

Chambers, T.J., Weir, R.C., Grakoui, A., McCourt, D.W., Bazan, J.F., Fletterick, R.J., and Rice, C.M. (1990b). Evidence That the N-Terminal Domain of Nonstructural Protein NS3 from Yellow-Fever Virus Is a Serine Protease Responsible for Site-Specific Cleavages in the Viral Polyprotein. *Proc Natl Acad Sci U S A* **87**, 8898-8902.

Chambers, T.J., Weir, R.C., Grakoui, A., McCourt, D.W., Bazan, J.F., Fletterick, R.J., and Rice, C.M. (1990c). Evidence that the N-terminal domain of nonstructural protein NS3 from yellow fever virus is a serine protease responsible for site-specific cleavages in the viral polyprotein. *Proc Natl Acad Sci U S A* **87**, 8898-8902.

Chan, Y.K., and Gack, M.U. (2016). A phosphomimetic-based mechanism of dengue virus to antagonize innate immunity. *Nature immunology* **17**, 523-530.

Chandramouli, S., Joseph, J.S., Daudenarde, S., Gatchalian, J., Cornillez-Ty, C., and Kuhn, P. (2010). Serotype-specific structural differences in the protease-cofactor complexes of the dengue virus family. *J Virol* **84**, 3059-3067.

Chanprapaph, S., Saparpakorn, P., Sangma, C., Niyomrattanakit, P., Hannongbua, S., Angsuthanasombat, C., and Katzenmeier, G. (2005). Competitive inhibition of the dengue virus NS3 serine protease by synthetic peptides representing polyprotein cleavage sites. *Biochem Biophys Res Commun* **330**, 1237-1246.

Chao, L.H., Klein, D.E., Schmidt, A.G., Pena, J.M., and Harrison, S.C. (2014). Sequential conformational rearrangements in flavivirus membrane fusion. *Elife* **3**.

Chappell, K.J., Nall, T.A., Stoermer, M.J., Fang, N.X., Tyndall, J.D., Fairlie, D.P., and Young, P.R. (2005). Site-directed mutagenesis and kinetic studies of the West Nile virus NS3 protease identify key enzyme-substrate interactions. *J Biol Chem* **280**, 2896-2903.

Chappell, K.J., Stoermer, M.J., Fairlie, D.P., and Young, P.R. (2006). Insights to substrate binding and processing by West Nile virus NS3 protease through combined modeling, protease mutagenesis, and kinetic studies. *J Biol Chem* **281**, 38448-38458.

Chappell, K.J., Stoermer, M.J., Fairlie, D.P., and Young, P.R. (2007). Generation and characterization of proteolytically active and highly stable truncated and full-length recombinant West Nile virus NS3. *Protein Expr Purif* **53**, 87-96.

Chappell, K.J., Stoermer, M.J., Fairlie, D.P., and Young, P.R. (2008). Mutagenesis of the West Nile virus NS2B cofactor domain reveals two regions essential for protease activity. *J Gen Virol* **89**, 1010-1014.

Che, P.L., Tang, H.L., and Li, Q.J. (2013). The interaction between claudin-1 and dengue viral prM/M protein for its entry. *Virology* **446**, 303-313.

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

Chen, H., Liu, L., Jones, S.A., Banavali, N., Kass, J., Li, Z., Zhang, J., Kramer, L.D., Ghosh, A.K., and Li, H. (2013). Selective inhibition of the West Nile virus methyltransferase by nucleoside analogs. *Antiviral research* **97**, 232-239.

Chen, S.T., Lin, Y.L., Huang, M.T., Wu, M.F., Cheng, S.C., Lei, H.Y., Lee, C.K., Chiou, T.W., Wong, C.H., and Hsieh, S.L. (2008). CLEC5A is critical for dengue-virus-induced lethal disease. *Nature* 453, 672-676.

Chen, S.T., Liu, R.S., Wu, M.F., Lin, Y.L., Chen, S.Y., Tan, D.T., Chou, T.Y., Tsai, I.S., Li, L., and Hsieh, S.L. (2012). CLEC5A regulates Japanese encephalitis virus-induced neuroinflammation and lethality. *PLoS Path* 8, e1002655.

Chen, X., Yang, K., Wu, C., Chen, C., Hu, C., Buzovetsky, O., Wang, Z., Ji, X., Xiong, Y., and Yang, H. (2016). Mechanisms of activation and inhibition of Zika virus NS2B-NS3 protease. *Cell Res* 26, 1260-1263.

Chen, Y., Maguire, T., Hileman, R.E., Fromm, J.R., Esko, J.D., Linhardt, R.J., and Marks, R.M. (1997b). Dengue virus infectivity depends on envelope protein binding to target cell heparan sulfate. *Nat Med* 3, 866-871.

Cheng, G., Cox, J., Wang, P., Krishnan, M.N., Dai, J., Qian, F., Anderson, J.F., and Fikrig, E. (2010). A C-type lectin collaborates with a CD45 phosphatase homolog to facilitate West Nile virus infection of mosquitoes. *Cell* 142, 714-725.

Cheng, Y., King, N.J., and Kesson, A.M. (2004). Major histocompatibility complex class I (MHC-I) induction by West Nile virus: involvement of 2 signaling pathways in MHC-I up-regulation. *The J Infect Dis* 189, 658-668.

Chernov, A.V., Shiryayev, S.A., Aleshin, A.E., Ratnikov, B.I., Smith, J.W., Liddington, R.C., and Strongin, A.Y. (2008). The two-component NS2B-NS3 proteinase represses DNA unwinding activity of the West Nile virus NS3 helicase. *J Biol Chem* 283, 17270-17278.

Chu, J.J., Leong, P.W., and Ng, M.L. (2006). Analysis of the endocytic pathway mediating the infectious entry of mosquito-borne flavivirus West Nile into *Aedes albopictus* mosquito (C6/36) cells. *Virology* 349, 463-475.

Chu, J.J., and Ng, M.L. (2004). Infectious entry of West Nile virus occurs through a clathrin-mediated endocytic pathway. *J Virol* 78, 10543-10555.

Chung, K.M., Liszewski, M.K., Nybakken, G., Davis, A.E., Townsend, R.R., Fremont, D.H., Atkinson, J.P., and Diamond, M.S. (2006). West Nile virus nonstructural protein NS1 inhibits complement activation by binding the regulatory protein factor H. *Proc Natl Acad Sci U S A* 103, 19111-19116.

Clum, S., Ebner, K.E., and Padmanabhan, R. (1997). Cotranslational membrane insertion of the serine proteinase precursor NS2B-NS3(Pro) of dengue virus type 2 is required for efficient in vitro processing and is mediated through the hydrophobic regions of NS2B. *J Biol Chem* 272, 30715-30723.

Coller, B.A.G., Clements, D.E., Bett, A.J., Sagar, S.L., and Ter Meulen, J.H. (2011). The development of recombinant subunit envelope-based against dengue virus induced disease. *Vaccine* 29, 7267-7275.

Coloma, J., Jain, R., Rajashankar, K.R., Garcia-Sastre, A., and Aggarwal, A.K. (2016). Structures of NS5 Methyltransferase from Zika Virus. *Cell Reports* 16, 3097-3102.

Colpitts, T.M., Barthel, S., Wang, P.H., and Fikrig, E. (2011). Dengue Virus Capsid Protein Binds Core Histones and Inhibits Nucleosome Formation in Human Liver Cells. *Plos One* 6.

Colpitts, T.M., Conway, M.J., Montgomery, R.R., and Fikrig, E. (2012). West Nile virus: biology, transmission, and human infection. *Clin Microbiol Rev* 25, 635-648.

Coutard, B., Decroly, E., Li, C., Sharff, A., Lescar, J., Bricogne, G., and Barral, K. (2014). Assessment of Dengue virus helicase and methyltransferase as targets for fragment-based drug discovery. *Antiviral Res* 106, 61-70.

D'Ortenzio, E., Matheron, S., Yazdanpanah, Y., de Lamballerie, X., Hubert, B., Piorkowski, G., Maquart, M., Descamps, D., Damond, F., and Leparac-Goffart, I. (2016). Evidence of Sexual Transmission of Zika Virus. *N Engl J Med* 374, 2195-2198.

Daffis, S., Szretter, K.J., Schriewer, J., Li, J., Youn, S., Errett, J., Lin, T.Y., Schneller, S., Zust, R., Dong, H., et al. (2010). 2'-O methylation of the viral mRNA cap evades host restriction by IFIT family members. *Nature* 468, 452-456.

Dai, L., Song, J., Lu, X., Deng, Y.Q., Musyoki, A.M., Cheng, H., Zhang, Y., Yuan, Y., Song, H., Haywood, J., *et al.* (2016). Structures of the Zika Virus Envelope Protein and Its Complex with a Flavivirus Broadly Protective Antibody. *Cell host & microbe* **19**, 696-704.

Dalrymple, N.A., Cimica, V., and Mackow, E.R. (2015). Dengue Virus NS Proteins Inhibit RIG-I/MAVS Signaling by Blocking TBK1/IRF3 Phosphorylation: Dengue Virus Serotype 1 NS4A Is a Unique Interferon-Regulating Virulence Determinant. *mBio* **6**, e00553-00515.

Das, S., Laxminarayana, S.V., Chandra, N., Ravi, V., and Desai, A. (2009). Heat shock protein 70 on Neuro2a cells is a putative receptor for Japanese encephalitis virus. *Virology* **385**, 47-57.

Dauter, Z., Dauter, M., and Dodson, E. (2002). Jolly SAD. *Acta Crystallogr D Biol Crystallogr* **58**, 494-506.

Davis, C.W., Nguyen, H.Y., Hanna, S.L., Sanchez, M.D., Doms, R.W., and Pierson, T.C. (2006). West Nile virus discriminates between DC-SIGN and DC-SIGNR for cellular attachment and infection. *J Virol* **80**, 1290-1301.

de la Cruz, L., Chen, W.N., Graham, B., and Otting, G. (2014). Binding mode of the activity-modulating C-terminal segment of NS2B to NS3 in the dengue virus NS2B-NS3 protease. *FEBS J* **281**, 1517-1533.

de la Cruz, L., Nguyen, T.H., Ozawa, K., Shin, J., Graham, B., Huber, T., and Otting, G. (2011). Binding of low molecular weight inhibitors promotes large conformational changes in the dengue virus NS2B-NS3 protease: fold analysis by pseudocontact shifts. *J Am Chem Soc* **133**, 19205-19215.

Dehan, P., Hamers, C., Letellier, C., Couvreur, B., Kerkhofs, P., and Pastoret, P.P. (2001). Molecular biology of bovine viral diarrhea virus: a review. *Ann Med Vet* **145**, 39-46.

Dejnirattisai, W., Webb, A.I., Chan, V., Jumnainsong, A., Davidson, A., Mongkolsapaya, J., and Screaton, G. (2011). Lectin switching during dengue virus infection. *The J Infect Dis* **203**, 1775-1783.

Deng, J., Li, N., Liu, H., Zuo, Z., Liew, O.W., Xu, W., Chen, G., Tong, X., Tang, W., Zhu, J., *et al.* (2012). Discovery of novel small molecule inhibitors of dengue viral NS2B-NS3 protease using virtual screening and scaffold hopping. *J Med Chem* **55**, 6278-6293.

Diamond, M.S., Roberts, T.G., Edgil, D., Lu, B., Ernst, J., and Harris, E. (2000). Modulation of Dengue virus infection in human cells by alpha, beta, and gamma interferons. *J Virol* **74**, 4957-4966.

Dick, G.W., Kitchen, S.F., and Haddow, A.J. (1952). Zika virus. I. Isolations and serological specificity. *Trans R Soc Trop Med Hyg* **46**, 509-520.

Diederichs, K., and Karplus, P.A. (1997). Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat Struct Biol* **4**, 269-275.

Dodson, E. (2003). Is it jolly SAD? *Acta Crystallogr D Biol Crystallogr* **59**, 1958-1965.

Dokland, T., Walsh, M., Mackenzie, J.M., Khromykh, A.A., Ee, K.H., and Wang, S. (2004a). West Nile virus core protein; tetramer structure and ribbon formation. *Structure* **12**, 1157-1163.

Dokland, T., Walsh, M., Mackenzie, J.M., Khromykh, A.A., Ee, K.H., and Wang, S.F. (2004b). West Nile virus core protein: Tetramer structure and ribbon formation. *Structure* **12**, 1157-1163.

Dong, H.P., Liu, L.H., Zou, G., Zhao, Y.W., Li, Z., Lim, S.P., Shi, P.Y., and Li, H.M. (2010). Structural and Functional Analyses of a Conserved Hydrophobic Pocket of Flavivirus Methyltransferase. *J Biol Chem* **285**, 32586-32595.

Dowd, K.A., Ko, S.Y., Morabito, K.M., Yang, E.S., Pelc, R.S., DeMaso, C.R., Castilho, L.R., Abbink, P., Boyd, M., Nityanandam, R., *et al.* (2016). Rapid development of a DNA vaccine for Zika virus. *Science* **354**, 237-240.

Droll, D.A., Krishna Murthy, H.M., and Chambers, T.J. (2000). Yellow fever virus NS2B-NS3 protease: charged-to-alanine mutagenesis and deletion analysis define regions important for protease complex formation and function. *Virology* **275**, 335-347.

Durbin, A.P., Kirkpatrick, B.D., Pierce, K.K., Elwood, D., Larsson, C.J., Lindow, J.C., Tibery, C., Sabundayo, B.P., Shaffer, D., Talaat, K.R., *et al.* (2013a). A Single Dose of Any of Four Different Live Attenuated Tetravalent Dengue Vaccines Is Safe and Immunogenic in Flavivirus-naïve Adults: A Randomized, Double-blind Clinical Trial. *J Infect Dis* **207**, 957-965.

Durbin, A.P., Wright, P.F., Cox, A., Kagucia, W., Elwood, D., Henderson, S., Wanionek, K., Speicher, J., Whitehead, S.S., and Pletnev, A.G. (2013b). The live attenuated chimeric vaccine rWN/DEN4 Delta 30 is well-tolerated and immunogenic in healthy flavivirus-naïve adult volunteers. *Vaccine* 31, 5772-5777.

Ealick, S.E. (2000). Advances in multiple wavelength anomalous diffraction crystallography. *Current Opinion in Chem Biol* 4, 495-499.

Egloff, M.P., Decroly, E., Malet, H., Selisko, B., Benarroch, D., Ferron, F., and Canard, B. (2007). Structural and functional analysis of methylation and 5'-RNA sequence requirements of short capped RNAs by the methyltransferase domain of dengue virus NS5. *J Mol Biol* 372, 723-736.

Eisenberg, D., Schwarz, E., Komaromy, M., and Wall, R. (1984). Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J Mol Biol* 179, 125-142.

Elong Ngono, A., and Shresta, S. (2018). Immune Response to Dengue and Zika. *Annu Rev Immunol*.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60, 2126-2132.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D* 66, 486-501.

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

Erez, E., Fass, D., and Bibi, E. (2009). How intramembrane proteases bury hydrolytic reactions in the membrane. *Nature* 459, 371-378.

Espada-Murao, L.A., and Morita, K. (2011). Delayed Cytosolic Exposure of Japanese Encephalitis Virus Double-Stranded RNA Impedes Interferon Activation and Enhances Viral Dissemination in Porcine Cells. *J Virol* 85, 6736-6749.

Evans, P.R. (2007). An introduction to stereochemical restraints. *Acta Crystallogr D Biol Crystallogr* 63, 58-61.

Evans, P.R. (2011). An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr D Biol Crystallogr* 67, 282-292.

Ezgimen, M., Lai, H., Mueller, N.H., Lee, K., Cuny, G., Ostrov, D.A., and Padmanabhan, R. (2012). Characterization of the 8-hydroxyquinoline scaffold for inhibitors of West Nile virus serine protease. *Antiviral Res* 94, 18-24.

Failla, C., Tomei, L., and Defrancesco, R. (1994). Both Ns3 and Ns4a Are Required for Proteolytic Processing of Hepatitis-C Virus Nonstructural Proteins. *J Virol* 68, 3753-3760.

Falgout, B., Pethel, M., Zhang, Y.M., and Lai, C.J. (1991). Both nonstructural proteins NS2B and NS3 are required for the proteolytic processing of dengue virus nonstructural proteins. *J Virol* 65, 2467-2475.

Faria, N.R., Azevedo, R., Kraemer, M.U.G., Souza, R., Cunha, M.S., Hill, S.C., Theze, J., Bonsall, M.B., Bowden, T.A., Rissanen, I., et al. (2016). Zika virus in the Americas: Early epidemiological and genetic findings. *Science* 352, 345-349.

Feng, Q., Hato, S.V., Langereis, M.A., Zoll, J., Virgen-Slane, R., Peisley, A., Hur, S., Semler, B.L., van Rij, R.P., and van Kuppeveld, F.J. (2012). MDA5 detects the double-stranded RNA replicative form in picornavirus-infected cells. *Cell Rep* 2, 1187-1196.

Flipse, J., Wilschut, J., and Smit, J.M. (2013). Molecular Mechanisms Involved in Antibody-Dependent Enhancement of Dengue Virus Infection in Humans. *Traffic* 14, 25-35.

Freire, J.M., Veiga, A.S., Conceicao, T.M., Kowalczyk, W., Mohana-Borges, R., Andreu, D., Santos, N.C., Da Poian, A.T., and Castanho, M.A.R.B. (2013). Intracellular Nucleic Acid Delivery by the Supercharged Dengue Virus Capsid Protein. *Plos One* 8.

Frieden, T.R., Schuchat, A., and Petersen, L.R. (2016). Zika Virus 6 Months Later. *JAMA* 316, 1443-1444.

Gallei, A., Blome, S., Gilgenbach, S., Tautz, N., Moennig, V., and Becher, P. (2008). Cytopathogenicity of classical swine fever virus correlates with attenuation in the natural host. *J Virol* 82, 9717-9729.

Ganesh, V.K., Muller, N., Judge, K., Luan, C.H., Padmanabhan, R., and Murthy, K.H. (2005). Identification and characterization of nonsubstrate based inhibitors of the essential dengue and West Nile virus proteases. *Bioorg Med Chem* 13, 257-264.

Gao, F., Duan, X.Q., Lu, X., Liu, Y.M., Zheng, L., Ding, Z.H., and Li, J. (2010). Novel binding between pre-membrane protein and claudin-1 is required for efficient dengue virus entry. *Biochem Biophys Res Comm* 391, 952-957.

Gebhard, L.G., Filomatori, C.V., and Gamarnik, A.V. (2011). Functional RNA elements in the dengue virus genome. *Viruses* 3, 1739-1756.

Geiss, B.J., Thompson, A.A., Andrews, A.J., Sons, R.L., Gari, H.H., Keenan, S.M., and Peersen, O.B. (2009). Analysis of Flavivirus NS5 Methyltransferase Cap Binding. *J Mol Bio* 385, 1643-1654.

Germi, R., Crance, J.M., Garin, D., Guimet, J., Lortat-Jacob, H., Ruigrok, R.W., Zarski, J.P., and Drouet, E. (2002). Heparan sulfate-mediated binding of infectious dengue virus type 2 and yellow fever virus. *Virology* 292, 162-168.

Gerold, G., Bruening, J., Weigel, B., and Pietschmann, T. (2017). Protein Interactions during the Flavivirus and Hepacivirus Life Cycle. *Mol Cell Proteomics* 16, S75-S91.

Gillespie, L.K., Hoenen, A., Morgan, G., and Mackenzie, J.M. (2010). The Endoplasmic Reticulum Provides the Membrane Platform for Biogenesis of the Flavivirus Replication Complex. *J Virol* 84, 10438-10447.

Go, Y.Y., Balasuriya, U.B., and Lee, C.K. (2014). Zoonotic encephalitides caused by arboviruses: transmission and epidemiology of alphaviruses and flaviviruses. *Clin Exp Vaccine Res* 3, 58-77.

Gouvea, I.E., Izidoro, M.A., Judice, W.A.S., Cezari, M.H.S., Caliendo, G., Santagada, V., dos Santos, C.N.D., Queiroz, M.H., Juliano, M.A., Young, P.R., *et al.* (2007). Substrate specificity of recombinant dengue 2 virus NS2B-NS3 protease: Influence of natural and unnatural basic amino acids on hydrolysis of synthetic fluorescent substrates. *Archives of Biochemistry and Biophysics* 457, 187-196.

Gradman, A.H., Schmieder, R.E., Lins, R.L., Nussberger, J., Chiang, Y., and Bedigian, M.P. (2005). Aliskiren, a novel orally effective renin inhibitor, provides dose-dependent antihypertensive efficacy and placebo-like tolerability in hypertensive patients. *Circulation* 111, 1012-1018.

Gransden, W.R. (1999). Topley and Wilson's Microbiology and Microbial Infections, 9th edition (CD-ROM). *J Clin Pathol* 52, 237-238.

Grant, A., Ponia, S.S., Tripathi, S., Balasubramaniam, V., Miorin, L., Sourisseau, M., Schwarz, M.C., Sanchez-Seco, M.P., Evans, M.J., Best, S.M., *et al.* (2016). Zika Virus Targets Human STAT2 to Inhibit Type I Interferon Signaling. *Cell host & microbe* 19, 882-890.

Gruba, N., Rodriguez Martinez, J.I., Grzywa, R., Wysocka, M., Skorenski, M., Burmistrz, M., Lecka, M., Lesner, A., Sienczyk, M., and Pyrc, K. (2016). Substrate profiling of Zika virus NS2B-NS3 protease. *FEBS Lett* 590, 3459-3468.

Grummer, B., Grotha, S., and Greiser-Wilke, I. (2004). Bovine viral diarrhoea virus is internalized by clathrin-dependent receptor-mediated endocytosis. *J Vet Med B* 51, 427-432.

Gubler, D.J., Vasilakis, N., and Musso, D. (2017). History and Emergence of Zika Virus. *J Infect Dis* 216, S860-S867.

Guedes, D.R., Paiva, M.H., Donato, M.M., Barbosa, P.P., Krokovsky, L., Rocha, S., Saraiva, K., Crespo, M.M., Rezende, T.M., Wallau, G.L., *et al.* (2017). Zika virus replication in the mosquito *Culex quinquefasciatus* in Brazil. *Emerg Microbes Infect* 6, e69.

Guo, J.T., Hayashi, J., and Seeger, C. (2005). West Nile virus inhibits the signal transduction pathway of alpha interferon. *J Virol* 79, 1343-1350.

Guy, B., Barrere, B., Malinowski, C., Saville, M., Teyssou, R., and Lang, J. (2011). From research to phase III: Preclinical, industrial and clinical development of the Sanofi Pasteur tetravalent dengue vaccine. *Vaccine* 29, 7229-7241.

Hahm, B., Han, D.S., Back, S.H., Song, O.K., Cho, M.J., Kim, C.J., Shimotohno, K., and Jang, S.K. (1995). NS3-4A of hepatitis C virus is a chymotrypsin-like protease. *J Virol* 69, 2534-2539.

Halstead, S.B. (2007). Dengue. *Lancet* 370, 1644-1652.

Halstead, S.B. (2014). Dengue Antibody- Dependent Enhancement: Knowns and Unknowns. *Microbiology Spectrum* 2.

Halstead, S.B., Suaya, J.A., and Shepard, D.S. (2007). The burden of dengue infection. *Lancet* 369, 1410-1411.

Hamel, R., Dejarnac, O., Wichit, S., Ekchariyawat, P., Neyret, A., Luplertlop, N., Perera-Lecoin, M., Surasombatpattana, P., Talignani, L., Thomas, F., *et al.* (2015). Biology of Zika Virus Infection in Human Skin Cells. *J Virol* 89, 8880-8896.

Hammamy, M.Z., Haase, C., Hammami, M., Hilgenfeld, R., and Steinmetzer, T. (2013). Development and characterization of new peptidomimetic inhibitors of the West Nile virus NS2B-NS3 protease. *ChemMedChem* 8, 231-241.

Han, D.S., Hahm, B., Rho, H.M., and Jang, S.K. (1995). Identification of the protease domain in NS3 of hepatitis C virus. *J Gen Virol* 76 (Pt 4), 985-993.

Hasan, S.S., Sevana, M., Kuhn, R.J., and Rossmann, M.G. (2018). Structural biology of Zika virus and other flaviviruses. *Nat Struct Mol Biol* 25, 13-20.

Hayward, S., and Berendsen, H.J. (1998). Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. *Proteins* 30, 144-154.

Hayward, S., and Lee, R.A. (2002). Improvements in the analysis of domain motions in proteins from conformational change: DynDom version 1.50. *J Mol Graph Model* 21, 181-183.

He, Z., Zhu, X., Wen, W., Yuan, J., Hu, Y., Chen, J., An, S., Dong, X., Lin, C., Yu, J., *et al.* (2016). Dengue Virus Subverts Host Innate Immunity by Targeting Adaptor Protein MAVS. *J Virol* 90, 7219-7230.

Headd, J.J., Echols, N., Afonine, P.V., Grosse-Kunstleve, R.W., Chen, V.B., Moriarty, N.W., Richardson, D.C., Richardson, J.S., and Adams, P.D. (2012). Use of knowledge-based restraints in phenix.refine to improve macromolecular refinement at low resolution. *Acta Crystallogr D Biol Crystallogr* 68, 381-390.

Hershey, A.D., and Chase, M. (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J Gen Physiol* 36, 39-56.

Hershkovitz, O., Rosental, B., Rosenberg, L.A., Navarro-Sanchez, M.E., Jivov, S., Zilka, A., Gershoni-Yahalom, O., Brient-Litzler, E., Bedouelle, H., Ho, J.W., *et al.* (2009). NKp44 Receptor Mediates Interaction of the Envelope Glycoproteins from the West Nile and Dengue Viruses with NK Cells. *J Immuno* 183, 2610-2621.

Hershkovitz, O., Zilka, A., Bar-Ilan, A., Abutbul, S., Davidson, A., Mazzon, M., Kummerer, B.M., Monsoengo, A., Jacobs, M., and Porgador, A. (2008). Dengue virus replicon expressing the nonstructural proteins suffices to enhance membrane expression of HLA class I and inhibit lysis by human NK cells. *J Virol* 82, 7666-7676.

Hills, S.L., and Phillips, D.C. (2009). Past, present, and future of Japanese encephalitis. *Emerg Infect Dis* 15, 1333.

Holbrook, M.R. (2017a). Historical Perspectives on Flavivirus Research. *Viruses* 9.

Holbrook, M.R. (2017b). Historical Perspectives on Flavivirus Research. *Viruses-Basel* 9.

Holcomb, J., Spellmon, N., Zhang, Y., Doughan, M., Li, C., and Yang, Z. (2017). Protein crystallization: Eluding the bottleneck of X-ray crystallography. *AIMS Biophys* 4, 557-575.

Holliday, D.C. (1880). Dengue or Dandy Fever. *Public Health Pap Rep* 6, 166-169.

Hsieh, S.C., Wu, Y.C., Zou, G., Nerurkar, V.R., Shi, P.Y., and Wang, W.K. (2014). Highly conserved residues in the helical domain of dengue virus type 1 precursor membrane protein are involved in assembly, precursor membrane (prM) protein cleavage, and entry. *J Biol Chem* 289, 33149-33160.

Hunsperger, E.A., Yoksan, S., Buchy, P., Nguyen, V.C., Sekaran, S.D., Enria, D.A., Vazquez, S., Cartozian, E., Pelegrino, J.L., Artsob, H., *et al.* (2014). Evaluation of Commercially Available Diagnostic Tests for the Detection of Dengue Virus NS1 Antigen and Anti-Dengue Virus IgM Antibody. *Plos Neglected Tropical Diseases* 8.

Inohara, Chamaillard, McDonald, C., and Nunez, G. (2005). NOD-LRR proteins: role in host-microbial interactions and inflammatory disease. *Annu Rev Biochem* 74, 355-383.

Ishikawa, H., and Barber, G.N. (2008). STING is an endoplasmic reticulum adaptor that facilitates innate immune signalling. *Nature* 455, 674-678.

Ishikawa, H., Ma, Z., and Barber, G.N. (2009). STING regulates intracellular DNA-mediated, type I interferon-dependent innate immunity. *Nature* 461, 788-792.

Ishikawa, T., Yamanaka, A., and Konishi, E. (2014). A review of successful flavivirus vaccines and the problems with those flaviviruses for which vaccines are not yet available. *Vaccine* 32, 1326-1337.

Issur, M., Geiss, B.J., Bougie, I., Picard-Jean, F., Despins, S., Mayette, J., Hobdey, S.E., and Bisailon, M. (2009). The flavivirus NS5 protein is a true RNA guanylyltransferase that catalyzes a two-step reaction to form the RNA cap structure. *RNA* 15, 2340-2350.

Ivashkiv, L.B., and Donlin, L.T. (2014). Regulation of type I interferon responses. *Nat Rev Immunol* 14, 36-49.

Jacobs, M.G., Robinson, P.J., Bletchly, C., Mackenzie, J.M., and Young, P.R. (2000). Dengue virus nonstructural protein 1 is expressed in a glycosyl-phosphatidylinositol-linked form that is capable of signal transduction. *Faseb Journal* 14, 1603-1610.

Jain, R., Butler, K.V., Coloma, J., Jin, J., and Aggarwal, A.K. (2017). Development of a S-adenosylmethionine analog that intrudes the RNA-cap binding site of Zika methyltransferase. *Scientific Reports* 7.

Jain, R., Coloma, J., Garcia-Sastre, A., and Aggarwal, A.K. (2016). Structure of the NS3 helicase from Zika virus. *Nat Struct Mol Biol* 23, 752-754.

Jan, L.R., Yang, C.S., Trent, D.W., Falgout, B., and Lai, C.J. (1995). Processing of Japanese encephalitis virus non-structural proteins: NS2B-NS3 complex and heterologous proteases. *J Gen Virol* 76 (Pt 3), 573-580.

Jemielity, S., Wang, J.J., Chan, Y.K., Ahmed, A.A., Li, W.H., Monahan, S., Bu, X., Farzan, M., Freeman, G.J., Umetsu, D.T., *et al.* (2013). TIM-family Proteins Promote Infection of Multiple Enveloped Viruses through Virion-associated Phosphatidylserine. *PLoS Path* 9.

Jentes, E.S., Pomeroy, G., Gershman, M.D., Hill, D.R., Lemarchand, J., Lewis, R.F., Staples, J.E., Tomori, O., Wilder-Smith, A., Monath, T.P., *et al.* (2011). The revised global yellow fever risk map and recommendations for vaccination, 2010: consensus of the Informal WHO Working Group on Geographic Risk for Yellow fever. *Lancet Infect Dis* 11, 622-632.

Ji, H., Fraser, C.S., Yu, Y.H., Leary, J., and Doudna, J.A. (2004). Coordinated assembly of human translation initiation complexes by the hepatitis C virus internal ribosome entry site RNA. *Proc Natl Acad Sci U S A* 101, 16990-16995.

Jiang, R., Ye, J., Zhu, B.B., Song, Y.F., Chen, H.C., and Cao, S.B. (2014). Roles of TLR3 and RIG-I in Mediating the Inflammatory Response in Mouse Microglia following Japanese Encephalitis Virus Infection. *J Immuno Research*.

Joachimski, A. (2009). High-throughput crystallography for structural genomics. *Current Opinion in Structural Biology* 19, 573-584.

Johnston, P.A., Phillips, J., Shun, T.Y., Shinde, S., Lazo, J.S., Huryn, D.M., Myers, M.C., Ratnikov, B.I., Smith, J.W., Su, Y., *et al.* (2007). HTS identifies novel and specific uncompetitive inhibitors of the two-component NS2B-NS3 proteinase of West Nile virus. *Assay Drug Dev Technol* 5, 737-750.

Jones, M., Davidson, A., Hibbert, L., Gruenwald, P., Schlaak, J., Ball, S., Foster, G.R., and Jacobs, M. (2005). Dengue virus inhibits alpha interferon signaling by reducing STAT2 expression. *J Virol* 79, 5414-5420.

Junaid, M., Chalayat, C., Sehgelmeble Torrejon, A., Angsuthanasombat, C., Shutava, I., Lapins, M., Wikberg, J.E., and Katzenmeier, G. (2012). Enzymatic analysis of recombinant Japanese encephalitis virus NS2B(H)-NS3pro protease with fluorogenic model peptide substrates. *PLoS One* 7, e36872.

Junjhon, J., Edwards, T.J., Utaipat, U., Bowman, V.D., Holdaway, H.A., Zhang, W., Keelapang, P., Puttikhunt, C., Perera, R., Chipman, P.R., *et al.* (2010). Influence of pr-M cleavage on the heterogeneity of extracellular dengue virus particles. *J Virol* 84, 8353-8358.

Junjhon, J., Lausumpao, M., Supasa, S., Noisakran, S., Songjaeng, A., Saraithong, P., Chaichoun, K., Utaipat, U., Keelapang, P., Kanjanahaluethai, A., *et al.* (2008). Differential modulation of prM

cleavage, extracellular particle distribution, and virus infectivity by conserved residues at nonfuran consensus positions of the dengue virus pr-M junction. *J Virol* 82, 10776-10791.

Jusoh, T.N.A.M., and Shueb, R.H. (2017). Performance Evaluation of Commercial Dengue Diagnostic Tests for Early Detection of Dengue in Clinical Samples. *J Trop Med*.

Kabsch, W. (2010a). Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr* 66, 133-144.

Kabsch, W. (2010b). Xds. *Acta Crystallogr D Biol Crystallogr* 66, 125-132.

Kanai, R., Kar, K., Anthony, K., Gould, L.H., Ledizet, M., Fikrig, E., Marasco, W.A., Koski, R.A., and Modis, Y. (2006). Crystal structure of West Nile virus envelope glycoprotein reveals viral surface epitopes. *J Virol* 80, 11000-11008.

Kang, C., Keller, T.H., and Luo, D. (2017). Zika Virus Protease: An Antiviral Drug Target. *Trends Microbiol* 25, 797-808.

Kanneganti, T.D. (2010). Central roles of NLRs and inflammasomes in viral infection. *Nat Rev Immunol* 10, 688-698.

Kapoor, M., Zhang, L., Ramachandra, M., Kusukawa, J., Ebner, K.E., and Padmanabhan, R. (1995). Association between NS3 and NS5 proteins of dengue virus type 2 in the putative RNA replicase is linked to differential phosphorylation of NS5. *J Biol Chem* 270, 19100-19106.

Kaufmann, B., Chipman, P.R., Holdaway, H.A., Johnson, S., Fremont, D.H., Kuhn, R.J., Diamond, M.S., and Rossmann, M.G. (2009). Capturing a flavivirus pre-fusion intermediate. *PLoS Pathog* 5, e1000672.

Kell, A.M., and Gale, M., Jr. (2015). RIG-I in RNA virus recognition. *Virology* 479-480, 110-121.

Kiianitsa, K., Solinger, J.A., and Heyer, W.D. (2003). NADH-coupled microplate photometric assay for kinetic studies of ATP-hydrolyzing enzymes with low and high specific activities. *Anal Biochem* 321, 266-271.

Kim, J.L., Morgenstern, K.A., Lin, C., Fox, T., Dwyer, M.D., Landro, J.A., Chambers, S.P., Markland, W., Lepre, C.A., O'Malley, E.T., *et al.* (1997). Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide (vol 87, pg 343, 1996). *Cell* 89, U13-U13.

Kim, S.Y., Zhao, J., Liu, X.Y., Fraser, K., Lin, L., Zhang, X., Zhang, F.M., Dordick, J.S., and Linhardt, R.J. (2017). Interaction of Zika Virus Envelope Protein with Glycosaminoglycans. *Biochemistry* 56, 1151-1162.

Kim, Y.M., Gayen, S., Kang, C., Joy, J., Huang, Q., Chen, A.S., Wee, J.L., Ang, M.J., Lim, H.A., Hung, A.W., *et al.* (2013). NMR analysis of a novel enzymatically active unlinked dengue NS2B-NS3 protease complex. *J Biol Chem* 288, 12891-12900.

Kirkpatrick, B.D., Durbin, A.P., Pierce, K.K., Carmolli, M.P., Tibery, C.M., Grier, P.L., Hynes, N., Diehl, S.A., Elwood, D., Jarvis, A.P., *et al.* (2015). Robust and Balanced Immune Responses to All 4 Dengue Virus Serotypes Following Administration of a Single Dose of a Live Attenuated Tetravalent Dengue Vaccine to Healthy, Flavivirus-Naïve Adults. *J Infect Dis* 212, 702-710.

Klema, V.J., Padmanabhan, R., and Choi, K.H. (2015). Flaviviral Replication Complex: Coordination between RNA Synthesis and 5'-RNA Capping. *Viruses-Basel* 7, 4640-4656.

Knapman, K. (2001). High-throughput protein crystallography - Large amounts of data, analyzed simultaneously, provide new insights into protein-ligand interactions, accelerating the pace of drug discovery. *Chemical Innovation* 31, 22-27.

Knehans, T., Schuller, A., Doan, D.N., Nacro, K., Hill, J., Guntert, P., Madhusudhan, M.S., Weil, T., and Vasudevan, S.G. (2011). Structure-guided fragment-based in silico drug design of dengue protease inhibitors. *J Comput-Aided Mol Des* 25, 263-274.

Knipe, D.M.H., P.M. (2007). *Fields Virology* (Philadelphia: Lippincott-Raven Publishers).

Koh-Stenta, X., Joy, J., Wang, S.F., Kwek, P.Z., Wee, J.L., Wan, K.F., Gayen, S., Chen, A.S., Kang, C., Lee, M.A., *et al.* (2015). Identification of covalent active site inhibitors of dengue virus protease. *Drug Des Devel Ther* 9, 6389-6399.

Konishi, E., and Mason, P.W. (1993). Proper Maturation of the Japanese Encephalitis-Virus Envelope Glycoprotein Requires Cosynthesis with the Premembrane Protein. *J Virol* 67, 1672-1675.

Kouretova, J., Hammamy, M.Z., Epp, A., Harges, K., Kallis, S., Zhang, L., Hilgenfeld, R., Bartenschlager, R., and Steinmetzer, T. (2017). Effects of NS2B-NS3 protease and furin inhibition on West Nile and Dengue virus replication. *J Enzyme Inhib Med Chem* 32, 712-721.

Krey, T., Thiel, H.J., and Rumenapf, T. (2005). Acid-resistant bovine pestivirus requires activation for pH-triggered fusion during entry. *J Virol* 79, 4191-4200.

Krissinel, E., and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* 60, 2256-2268.

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372, 774-797.

Kroschewski, H., Allison, S.L., Heinz, F.X., and Mandl, C.W. (2003). Role of heparan sulfate for attachment and entry of tick-borne encephalitis virus. *Virology* 308, 92-100.

Kuadkitkan, A., Wikan, N., Fongsaran, C., and Smith, D.R. (2010). Identification and characterization of prohibitin as a receptor protein mediating DENV-2 entry into insect cells. *Virology* 406, 149-161.

Kuhn, P., Wilson, K., Patch, M.G., and Stevens, R.C. (2002a). The genesis of high-throughput structure-based drug discovery using protein crystallography. *Current Opinion in Chem Biol* 6, 704-710.

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

Kumar, A., Hou, S., Airo, A.M., Limonta, D., Mancinelli, V., Branton, W., Power, C., and Hobman, T.C. (2016). Zika virus inhibits type-I interferon production and downstream signaling. *EMBO reports* 17, 1766-1775.

Kumar, M., Roe, K., Orillo, B., Muruve, D.A., Nerurkar, V.R., Gale, M., Jr., and Verma, S. (2013). Inflammasome adaptor protein Apoptosis-associated speck-like protein containing CARD (ASC) is critical for the immune response and survival in West Nile virus encephalitis. *J Virol* 87, 3655-3667.

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

Kuno, G., Chang, G.J., Tsuchiya, K.R., Karabatsos, N., and Cropp, C.B. (1998). Phylogeny of the genus *Flavivirus*. *J Virol* 72, 73-83.

Kurosu, T., Chaichana, P., Yamate, M., Anantapreecha, S., and Ikuta, K. (2007). Secreted complement regulatory protein clusterin interacts with dengue virus nonstructural protein 1. *Biochem Biophys Res Comm* 362, 1051-1056.

Lai, H., Sridhar Prasad, G., and Padmanabhan, R. (2013). Characterization of 8-hydroxyquinoline derivatives containing aminobenzothiazole as inhibitors of dengue virus type 2 protease in vitro. *Antiviral Res* 97, 74-80.

Larocca, R.A., Abbink, P., Peron, J.P.S., Zanotto, P.M.D., Iampietro, M.J., Badamchi-Zadeh, A., Boyd, M., Ng'ang'a, D., Kirilova, M., Nityanandam, R., *et al.* (2016). Vaccine protection against Zika virus from Brazil. *Nature* 536, 474-+.

Leary, T.P., Muerhoff, A.S., Simons, J.N., Pilot-Matias, T.J., Erker, J.C., Chalmers, M.L., Schlauder, G.G., Dawson, G.J., Desai, S.M., and Mushahwar, I.K. (1996). Sequence and genomic organization of GBV-C: a novel member of the flaviviridae associated with human non-A-E hepatitis. *J Med Virol* 48, 60-67.

Ledgerwood, J.E., Pierson, T.C., Hubka, S.A., Desai, N., Rucker, S., Gordon, I.J., Enama, M.E., Nelson, S., Nason, M., Gu, W.J., *et al.* (2011). A West Nile virus DNA Vaccine Utilizing a Modified Promoter Induces Neutralizing Antibody in Younger and Older Healthy Adults in a Phase I Clinical Trial. *J Infect Dis* 203, 1396-1404.

Lee, C.J., Liao, C.L., and Lin, Y.L. (2005). Flavivirus activates phosphatidylinositol 3-kinase signaling to block caspase-dependent apoptotic cell death at the early stage of virus infection. *J Virol* 79, 8388-8399.

Lee, E., Pavy, M., Young, N., Freeman, C., and Lobigs, M. (2006). Antiviral effect of the heparan sulfate mimetic, PI-88, against dengue and encephalitic flaviviruses. *Antiviral research* 69, 31-38.

Lee, H., Ren, J., Nocadello, S., Rice, A.J., Ojeda, I., Light, S., Minasov, G., Vargas, J., Nagarathnam, D., Anderson, W.F., *et al.* (2016). Identification of novel small molecule inhibitors against NS2B/NS3 serine protease from Zika virus. *Antiviral Res* 139, 49-58.

Lehrer, A.T., and Holbrook, M.R. (2011). Tick-borne Encephalitis Vaccines. *J Bioterror Biodef* 2011, 3.

Lei, J., Hansen, G., Nitsche, C., Klein, C.D., Zhang, L., and Hilgenfeld, R. (2016). Crystal structure of Zika virus NS2B-NS3 protease in complex with a boronate inhibitor. *Science*.

Leppard, K.N.D.A. (2007). *Introduction to Modern Virology* (Blackwell Publishing Limited).

Lescar, J., Lim, S.P., Shi, S., (2012). *Molecular Virology and Control of Flaviviruses* (Caister Academic Press).

Lescar, J., Luo, D., Xu, T., Sampath, A., Lim, S.P., Canard, B., and Vasudevan, S.G. (2008). Towards the design of antiviral inhibitors against flaviviruses: the case for the multifunctional NS3 protein from Dengue virus as a target. *Antiviral Res* 80, 94-101.

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

Li, H., Clum, S., You, S., Ebner, K.E., and Padmanabhan, R. (1999). The serine protease and RNA-stimulated nucleoside triphosphatase and RNA helicase functional domains of dengue virus type 2 NS3 converge within a region of 20 amino acids. *J Virol* 73, 3108-3116.

Li, J., Lim, S.P., Beer, D., Patel, V., Wen, D., Tumanut, C., Tully, D.C., Williams, J.A., Jiricek, J., Priestley, J.P., *et al.* (2005). Functional profiling of recombinant NS3 proteases from all four serotypes of dengue virus using tetrapeptide and octapeptide substrate libraries. *J Biol Chem* 280, 28766-28774.

Li, L., Lok, S.M., Yu, I.M., Zhang, Y., Kuhn, R.J., Chen, J., and Rossmann, M.G. (2008a). The flavivirus precursor membrane-envelope protein complex: Structure and maturation. *Science* 319, 1830-1834.

Li, Y., Li, Q., Wong, Y.L., Liew, L.S., and Kang, C. (2015). Membrane topology of NS2B of dengue virus revealed by NMR spectroscopy. *Biochim Biophys Acta* 1848, 2244-2252.

Li, Y., Phoo, W.W., Loh, Y.R., Zhang, Z., Ng, E.Y., Wang, W., Keller, T.H., Luo, D., and Kang, C. (2017a). Structural characterization of the linked NS2B-NS3 protease of Zika virus. *FEBS Lett* 591, 2338-2347.

Li, Y., Zhang, Z., Phoo, W.W., Loh, Y.R., Wang, W., Liu, S., Chen, M.W., Hung, A.W., Keller, T.H., Luo, D., *et al.* (2017b). Structural Dynamics of Zika Virus NS2B-NS3 Protease Binding to Dipeptide Inhibitors. *Structure* 25, 1242-1250 e1243.

Li, Y.J., Kakinami, C., Li, Q., Yang, B.J., and Li, H.W. (2013). Human Apolipoprotein A-I Is Associated with Dengue Virus and Enhances Virus Infection through SR-BI. *Plos One* 8.

Li, Z., Brecher, M., Deng, Y.Q., Zhang, J., Sakamuru, S., Liu, B.B., Huang, R.L., Koetzner, C.A., Allen, C.A., Jones, S.A., *et al.* (2017c). Existing drugs as broad-spectrum and potent inhibitors for Zika virus by targeting NS2B-NS3 interaction. *Cell Research* 27, 1046-1064.

Li, Z., Khaliq, M., Zhou, Z., Post, C.B., Kuhn, R.J., and Cushman, M. (2008b). Design, synthesis, and biological evaluation of antiviral agents targeting flavivirus envelope proteins. *J Med Chem* 51, 4660-4671.

Lieberman, M.M., Nerurkar, V.R., Luo, H.Y., Cropp, B., Carrion, R., de la Garza, M., Coller, B.A., Clements, D., Ogata, S., Wong, T., *et al.* (2009). Immunogenicity and Protective Efficacy of a Recombinant Subunit West Nile virus Vaccine in Rhesus Monkeys. *Clinical and Vaccine Immunology* 16, 1332-1337.

Lim, H.A., Ang, M.J., Joy, J., Poulsen, A., Wu, W., Ching, S.C., Hill, J., and Chia, C.S. (2013). Novel agmatine dipeptide inhibitors against the West Nile virus NS2B/NS3 protease: a P3 and N-cap optimization study. *Eur J Med Chem* 62, 199-205.

Lim, S.P., Sonntag, L.S., Noble, C., Nilar, S.H., Ng, R.H., Zou, G., Monaghan, P., Chung, K.Y., Dong, H.P., Liu, B.P., *et al.* (2011). Small Molecule Inhibitors That Selectively Block Dengue Virus Methyltransferase. *J Biol Chem* 286, 6233-6240.

Lin, C., Lindenbach, B.D., Pragai, B.M., McCourt, D.W., and Rice, C.M. (1994). Processing in the hepatitis C virus E2-NS2 region: identification of p7 and two distinct E2-specific products with different C termini. *J Virol* 68, 5063-5073.

Lin, H.H., Yip, B.S., Huang, L.M., and Wu, S.C. (2018). Zika virus structural biology and progress in vaccine development. *Biotechnol Adv* 36, 47-53.

Lin, K.C., Chang, H.L., and Chang, R.Y. (2004a). Accumulation of a 3'-terminal genome fragment in Japanese encephalitis virus-infected mammalian and mosquito cells. *J Virol* 78, 5133-5138.

Lin, R.J., Liao, C.L., Lin, E., and Lin, Y.L. (2004b). Blocking of the alpha interferon-induced Jak-Stat signaling pathway by Japanese encephalitis virus infection. *J Virol* 78, 9285-9294.

Lindenbach, B.D. (2013). Virion assembly and release. *Curr Top Microbiol Immunol* 369, 199-218.

Lindenbach, B.D., and Rice, C.M. (2003a). Evasive maneuvers by hepatitis C virus. *Hepatology* 38, 769-771.

Lindenbach, B.D., and Rice, C.M. (2003b). Molecular biology of flaviviruses. *Adv Virus Res* 59, 23-61.

Lindenbach, B.D., and Rice, C.M. (2005). Unravelling hepatitis C virus replication from genome to function. *Nature* 436, 933-938.

Liu, H.M., Loo, Y.M., Horner, S.M., Zornetzer, G.A., Katze, M.G., and Gale, M., Jr. (2012). The mitochondrial targeting chaperone 14-3-3epsilon regulates a RIG-I translocon that mediates membrane association and innate antiviral immunity. *Cell host & microbe* 11, 528-537.

Liu, W.J., Chen, H.B., Wang, X.J., Huang, H., and Khromykh, A.A. (2004). Analysis of adaptive mutations in Kunjin virus replicon RNA reveals a novel role for the flavivirus nonstructural protein NS2A in inhibition of beta interferon promoter-driven transcription. *J Virol* 78, 12225-12235.

Liu, W.J., Wang, X.J., Clark, D.C., Lobigs, M., Hall, R.A., and Khromykh, A.A. (2006). A single amino acid substitution in the West Nile virus nonstructural protein NS2A disables its ability to inhibit alpha/beta interferon induction and attenuates virus virulence in mice. *J Virol* 80, 2396-2404.

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. *J Virol* 79, 1934-1942.

Liu, Z., Zhou, T., Lai, Z., Zhang, Z., Jia, Z., Zhou, G., Williams, T., Xu, J., Gu, J., Zhou, X., *et al.* (2017). Competence of *Aedes aegypti*, *Ae. albopictus*, and *Culex quinquefasciatus* Mosquitoes as Zika Virus Vectors, China. *Emerg Infect Dis* 23, 1085-1091.

Lo, M.C., Aulabaugh, A., Jin, G., Cowling, R., Bard, J., Malamas, M., and Ellestad, G. (2004). Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Analytical biochemistry* 332, 153-159.

Lobigs, M., Mullbacher, A., and Lee, E. (2004). Evidence that a mechanism for efficient flavivirus budding upregulates MHC class I. *Immunology and cell biology* 82, 184-188.

Lodeiro, M.F., Filomatori, C.V., and Gamarnik, A.V. (2009). Structural and functional studies of the promoter element for dengue virus RNA replication. *J Virol* 83, 993-1008.

Lohr, K., Knox, J.E., Phong, W.Y., Ma, N.L., Yin, Z., Sampath, A., Patel, S.J., Wang, W.L., Chan, W.L., Rao, K.R., *et al.* (2007). Yellow fever virus NS3 protease: peptide-inhibition studies. *J Gen Virol* 88, 2223-2227.

Lok, S.M., Kostyuchenko, V., Nybakken, G.E., Holdaway, H.A., Battisti, A.J., Sukupolvi-Petty, S., Sedlak, D., Fremont, D.H., Chipman, P.R., Roehrig, J.T., *et al.* (2008). Binding of a neutralizing antibody to dengue virus alters the arrangement of surface glycoproteins. *Nat Struct Mol Biol* 15, 312-317.

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. *J Virol* 76, 5480-5491.

Low, J.G., Ooi, E.E., and Vasudevan, S.G. (2017). Current Status of Dengue Therapeutics Research and Development. *J Infect Dis* 215, S96-S102.

Luo, D., Vasudevan, S.G., and Lescar, J. (2015). The flavivirus NS2B-NS3 protease-helicase as a target for antiviral drug development. *Antiviral Res* 118, 148-158.

Luo, D., Wei, N., Doan, D.N., Paradkar, P.N., Chong, Y., Davidson, A.D., Kotaka, M., Lescar, J., and Vasudevan, S.G. (2010). Flexibility between the protease and helicase domains of the dengue virus NS3 protein conferred by the linker region and its functional implications. *J Biol Chem* 285, 18817-18827.

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

Luo, D., Xu, T., Watson, R.P., Scherer-Becker, D., Sampath, A., Jahnke, W., Yeong, S.S., Wang, C.H., Lim, S.P., Strongin, A., *et al.* (2008b). Insights into RNA unwinding and ATP hydrolysis by the flavivirus NS3 protein. *EMBO J* 27, 3209-3219.

Ma, L., Jones, C.T., Groesch, T.D., Kuhn, R.J., and Post, C.B. (2004). Solution structure of dengue virus capsid protein reveals another fold. *Proc Natl Acad Sci U S A* 101, 3414-3419.

Ma, Y., Yates, J., Liang, Y., Lemon, S.M., and Yi, M. (2008). NS3 helicase domains involved in infectious intracellular hepatitis C virus particle assembly. *J Virol* 82, 7624-7639.

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., and Westaway, E.G. (2001). Assembly and maturation of the flavivirus Kunjin virus appear to occur in the rough endoplasmic reticulum and along the secretory pathway, respectively. *J Virol* 75, 10787-10799.

Makino, Y., Tadano, M., Anzai, T., Ma, S.P., Yasuda, S., and Fukunaga, T. (1989). Detection of Dengue-4 Virus Core Protein in the Nucleus .2. Antibody against Dengue-4 Core Protein Produced by a Recombinant Baculovirus Reacts with the Antigen in the Nucleus. *J Gen Virol* 70, 1417-1425.

Manoff, S.B., George, S.L., Bett, A.J., Yelmene, M.L., Dhanasekaran, G., Eggemeyer, L., Sausser, M.L., Dubey, S.A., Casimiro, D.R., Clements, D.E., *et al.* (2015). Preclinical and clinical development of a dengue recombinant subunit vaccine. *Vaccine* 33, 7126-7134.

Manokaran, G., Finol, E., Wang, C., Gunaratne, J., Bahl, J., Ong, E.Z., Tan, H.C., Sessions, O.M., Ward, A.M., Gubler, D.J., *et al.* (2015). Dengue subgenomic RNA binds TRIM25 to inhibit interferon expression for epidemiological fitness. *Science* 350, 217-221.

Mastrangelo, E., Milani, M., Bollati, M., Selisko, B., Peyrane, F., Pandini, V., Sorrentino, G., Canard, B., Konarev, P.V., Svergun, D.I., *et al.* (2007). Crystal structure and activity of Kunjin virus NS3 helicase; protease and helicase domain assembly in the full length NS3 protein. *J Mol Bio* 372, 444-455.

Matusan, A.E., Pryor, M.J., Davidson, A.D., and Wright, P.J. (2001). Mutagenesis of the Dengue virus type 2 NS3 protein within and outside helicase motifs: effects on enzyme activity and virus replication. *J Virol* 75, 9633-9643.

Mazzon, M., Jones, M., Davidson, A., Chain, B., and Jacobs, M. (2009). Dengue virus NS5 inhibits interferon-alpha signaling by blocking signal transducer and activator of transcription 2 phosphorylation. *The J Infect Dis* 200, 1261-1270.

Mc, L.D. (1955). Multiplication of viruses in mosquitoes following feeding and injection into the body cavity. *Aust J Exp Biol Med Sci* 33, 53-65.

McCoy, A.J. (2007). Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr D Biol Crystallogr* 63, 32-41.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007b). Phaser crystallographic software. *J Appl Crystallogr* 40, 658-674.

Medigeshi, G.R., Lancaster, A.M., Hirsch, A.J., Briesse, T., Lipkin, W.I., Defilippis, V., Fruh, K., Mason, P.W., Nikolich-Zugich, J., and Nelson, J.A. (2007). West Nile virus infection activates the unfolded protein response, leading to CHOP induction and apoptosis. *J Virol* 81, 10849-10860.

Meertens, L., Labeau, A., Dejarnac, O., Cipriani, S., Sinigaglia, L., Bonnet-Madin, L., Le Charpentier, T., Hafirassou, M.L., Zamborlini, A., Cao-Lormeau, V.M., *et al.* (2017). Axl Mediates ZIKA Virus Entry in Human Glial Cells and Modulates Innate Immune Responses. *Cell Reports* 18, 324-333.

Merrill, M.H., and Tenbroeck, C. (1935). The Transmission of Equine Encephalomyelitis Virus by *Aedes Aegypti*. *J Exp Med* 62, 687-695.

Minor, W., Cymborowski, M., Otwinowski, Z., and Chruszcz, M. (2006). HKL-3000: the integration of data reduction and structure solution - from diffraction images to an initial model in minutes. *Acta Crystallogr D Biol Crystallogr* 62, 859-866.

Miorin, L., Maestre, A.M., Fernandez-Sesma, A., and Garcia-Sastre, A. (2017). Antagonism of type I interferon by flaviviruses. *Biochem Biophys Res Comm* 492, 587-596.

Miorin, L., Romero-Brey, I., Maiuri, P., Hoppe, S., Krijnse-Locker, J., Bartenschlager, R., and Marcello, A. (2013). Three-Dimensional Architecture of Tick-Borne Encephalitis Virus Replication Sites and Trafficking of the Replicated RNA. *J Virol* 87, 6469-6481.

Mlakar, J., Korva, M., Tul, N., Popovic, M., Poljsak-Prijatelj, M., Mraz, J., Kolenc, M., Resman Rus, K., Vesnaver Vipotnik, T., Fabjan Vodusek, V., *et al.* (2016). Zika Virus Associated with Microcephaly. *N Engl J Med* 374, 951-958.

Modis, Y., Ogata, S., Clements, D., and Harrison, S.C. (2003). A ligand-binding pocket in the dengue virus envelope glycoprotein. *Proc Natl Acad Sci U S A* 100, 6986-6991.

Modis, Y., Ogata, S., Clements, D., and Harrison, S.C. (2004). Structure of the dengue virus envelope protein after membrane fusion. *Nature* 427, 313-319.

Monath, T.P., and Barrett, A.D.T. (2003). Pathogenesis and pathophysiology of yellow fever. *Flaviviruses: Pathogenesis and Immunity* 60, 343-395.

Monath, T.P., Liu, J., Kanesa-Thasan, N., Myers, G.A., Nichols, R., Deary, A., McCarthy, K., Johnson, C., Ermak, T., Shin, S., *et al.* (2006). A live, attenuated recombinant West Nile virus vaccine. *Proc Natl Acad Sci U S A* 103, 6694-6699.

Morabito, K.M., and Graham, B.S. (2017). Zika Virus Vaccine Development. *J Infect Dis* 216, S957-S963.

Moradpour, D., Penin, F., and Rice, C.M. (2007). Replication of hepatitis C virus. *Nat Rev Microbiol* 5, 453-463.

Morrison, J., Laurent-Rolle, M., Maestre, A.M., Rajsbaum, R., Pisanelli, G., Simon, V., Mulder, L.C., Fernandez-Sesma, A., and Garcia-Sastre, A. (2013). Dengue virus co-opts UBR4 to degrade STAT2 and antagonize type I interferon signaling. *PLoS Path* 9, e1003265.

Mueller, N.H., Yon, C., Ganesh, V.K., and Padmanabhan, R. (2007). Characterization of the West Nile virus protease substrate specificity and inhibitors. *Int J Biochem Cell Biol* 39, 606-614.

Muerhoff, A.S., Leary, T.P., Simons, J.N., Pilot-Matias, T.J., Dawson, G.J., Erker, J.C., Chalmers, M.L., Schlauder, G.G., Desai, S.M., and Mushahwar, I.K. (1995). Genomic organization of GB viruses A and B: two new members of the Flaviviridae associated with GB agent hepatitis. *J Virol* 69, 5621-5630.

Mukhopadhyay, S., Kuhn, R.J., and Rossmann, M.G. (2005). A structural perspective of the flavivirus life cycle. *Nat Rev Microbiol* 3, 13-22.

Muller, D.A., and Young, P.R. (2013). The flavivirus NS1 protein: molecular and structural biology, immunology, role in pathogenesis and application as a diagnostic biomarker. *Antiviral Res* 98, 192-208.

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. *J Virol* 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. *Proc Natl Acad Sci U S A* 100, 14333-14338.

Murphy, F.A., Fauquet, C.M., Bishop, D.H., Ghabrial, S.A., Jarvis, A.W., Martelli, G.P., Mayo, M.A., and Summers, M.D. (2012). *Virus taxonomy: classification and nomenclature of viruses*, Vol 10 (Springer Science & Business Media).

Murray, J.M., Aaskov, J.G., and Wright, P.J. (1993). Processing of the dengue virus type 2 proteins prM and C-prM. *J Gen Virol* 74 (Pt 2), 175-182.

Musso, D., Roche, C., Robin, E., Nhan, T., Teissier, A., and Cao-Lormeau, V.M. (2015). Potential sexual transmission of Zika virus. *Emerg Infect Dis* 21, 359-361.

Muthumani, K., Griffin, B.D., Agarwal, S., Kudchodkar, S.B., Reuschel, E.L., Choi, H., Kraynyak, K.A., Duperret, E.K., Keaton, A.A., Chung, C., *et al.* (2016). In vivo protection against ZIKV infection and pathogenesis through passive antibody transfer and active immunisation with a prMEnv DNA vaccine. *Npj Vaccines* 1.

Nall, T.A., Chappell, K.J., Stoermer, M.J., Fang, N.X., Tyndall, J.D., Young, P.R., and Fairlie, D.P. (2004). Enzymatic characterization and homology model of a catalytically active recombinant West Nile virus NS3 protease. *J Biol Chem* 279, 48535-48542.

Nasirudeen, A.M., Wong, H.H., Thien, P., Xu, S., Lam, K.P., and Liu, D.X. (2011). RIG-I, MDA5 and TLR3 synergistically play an important role in restriction of dengue virus infection. *PLoS Negl Trop Dis* 5, e926.

Nazmi, A., Mukherjee, S., Kundu, K., Dutta, K., Mahadevan, A., Shankar, S.K., and Basu, A. (2014). TLR7 is a key regulator of innate immunity against Japanese encephalitis virus infection. *Neurobiology of Disease* 69, 235-247.

Neill, J.D. (2013). o Molecular biology of bovine viral diarrhea virus. *Biologicals* 41, 2-7.

Newman, J., Egan, D., Walter, T.S., Meged, R., Berry, I., Ben Jelloul, M., Sussman, J.L., Stuart, D.I., and Perrakis, A. (2005). Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG plus strategy. *Acta Crystallogr D Biol Crystallogr* 61, 1426-1431.

Ng, W.C., Soto-Acosta, R., Bradrick, S.S., Garcia-Blanco, M.A., and Ooi, E.E. (2017). The 5' and 3' Untranslated Regions of the Flaviviral Genome. *Viruses* 9.

Nguyen, T.T.H., Lee, S., Wang, H.K., Chen, H.Y., Wu, Y.T., Lin, S.C., Kim, D.W., and Kim, D. (2013). In Vitro Evaluation of Novel Inhibitors against the NS2B-NS3 Protease of Dengue Fever Virus Type 4. *Molecules* 18, 15600-15612.

Nitsche, C., Behnam, M.A., Steuer, C., and Klein, C.D. (2012). Retro peptide-hybrids as selective inhibitors of the Dengue virus NS2B-NS3 protease. *Antiviral Res* 94, 72-79.

Nitsche, C., Schreier, V.N., Behnam, M.A., Kumar, A., Bartenschlager, R., and Klein, C.D. (2013). Thiazolidinone-peptide hybrids as dengue virus protease inhibitors with antiviral activity in cell culture. *J Med Chem* 56, 8389-8403.

Nitsche, C., Zhang, L., Weigel, L.F., Schilz, J., Graf, D., Bartenschlager, R., Hilgenfeld, R., and Klein, C.D. (2017). Peptide-Boronic Acid Inhibitors of Flaviviral Proteases: Medicinal Chemistry and Structural Biology. *J Med Chem* 60, 511-516.

Niyomrattanakit, P., Winoyanu Wattikun, P., Chanprapaph, S., Angsuthanasombat, C., Panyim, S., and Katzenmeier, G. (2004). Identification of residues in the dengue virus type 2 NS2B cofactor that are critical for NS3 protease activation. *J Virol* 78, 13708-13716.

Niyomrattanakit, P., Yahorava, S., Mutule, I., Mutulis, F., Petrovska, R., Prusis, P., Katzenmeier, G., and Wikberg, J.E. (2005). Dengue virus NS3 protease: studies on substrate specificity by using internally quenched synthetic peptides. *FEBS Journal* 272, 144-145.

Niyomrattanakit, P., Yahorava, S., Mutule, I., Mutulis, F., Petrovska, R., Prusis, P., Katzenmeier, G., and Wikberg, J.E.S. (2006). Probing the substrate specificity of the dengue virus type 2 NS3 serine protease by using internally quenched fluorescent peptides. *Biochem Journal* 397, 203-211.

Noble, C.G., Seh, C.C., Chao, A.T., and Shi, P.Y. (2012). Ligand-bound structures of the dengue virus protease reveal the active conformation. *J Virol* 86, 438-446.

Noble, C.G., and Shi, P.Y. (2012). Structural biology of dengue virus enzymes: towards rational design of therapeutics. *Antiviral Res* 96, 115-126.

Nybakken, G.E., Nelson, C.A., Chen, B.R., Diamond, M.S., and Fremont, D.H. (2006). Crystal structure of the West Nile virus envelope glycoprotein. *J Virol* 80, 11467-11474.

Overby, A.K., Popov, V.L., Niedrig, M., and Weber, F. (2010). Tick-Borne Encephalitis Virus Delays Interferon Induction and Hides Its Double-Stranded RNA in Intracellular Membrane Vesicles. *J Virol* 84, 8470-8483.

Pal, S., Dauner, A.L., Valks, A., Forshey, B.M., Long, K.C., Thaisomboonsuk, B., Sierra, G., Picos, V., Talmage, S., Morrison, A.C., *et al.* (2015). Multicountry Prospective Clinical Evaluation of Two Enzyme-Linked Immunosorbent Assays and Two Rapid Diagnostic Tests for Diagnosing Dengue Fever. *J Clin Microbiol* 53, 1092-1102.

Pambudi, S., Kawashita, N., Phanthanawiboon, S., Omokoko, M.D., Masrinoul, P., Yamashita, A., Limkittikul, K., Yasunaga, T., Takagi, T., Ikuta, K., *et al.* (2013). A small compound targeting the interaction between nonstructural proteins 2B and 3 inhibits dengue virus replication. *Biochem Biophys Res Commun* 440, 393-398.

Paul, D., and Bartenschlager, R. (2015). Flaviviridae Replication Organelles: Oh, What a Tangled Web We Weave. *Annual Review of Virology*, Vol 2 2, 289-310.

Paul, D., Madan, V., and Bartenschlager, R. (2014). Hepatitis C virus RNA replication and assembly: living on the fat of the land. *Cell Host Microbe* 16, 569-579.

Peeling, R.W., Artsob, H., Pelegrino, J.L., Buchy, P., Cardoso, M.J., Devi, S., Enria, D.A., Jeremy, F., Gubler, D.J., Guzman, M.C., *et al.* (2010). Evaluation of diagnostic tests: dengue. *Nat Rev Microbiol*, S30-S37.

Perera-Lecoin, M., Meertens, L., Carnec, X., and Amara, A. (2014). Flavivirus Entry Receptors: An Update. *Viruses-Basel* 6, 69-88.

Perera, R., Khaliq, M., and Kuhn, R.J. (2008). Closing the door on flaviviruses: entry as a target for antiviral drug design. *Antiviral Res* 80, 11-22.

Perera, R., and Kuhn, R.J. (2008). Structural proteomics of dengue virus. *Curr Opin Microbiol* 11, 369-377.

Perera, R., Riley, C., Isaac, G., Hopf-Jannasch, A.S., Moore, R.J., Weitz, K.W., Pasa-Tolic, L., Metz, T.O., Adamec, J., and Kuhn, R.J. (2012). Dengue virus infection perturbs lipid homeostasis in infected mosquito cells. *PLoS Pathog* 8, e1002584.

Peterhans, E., Bachofen, C., Stalder, H., and Schweizer, M. (2010). Cytopathic bovine viral diarrhea viruses (BVDV): emerging pestiviruses doomed to extinction. *Vet Res* 41.

Peterhans, E., and Schweizer, M. (2010). Pestiviruses: How to outmaneuver your hosts. *Vet Microbiol* 142, 18-25.

Petersen, E.E., Meaney-Delman, D., Neblett-Fanfair, R., Havers, F., Oduyebo, T., Hills, S.L., Rabe, I.B., Lambert, A., Abercrombie, J., Martin, S.W., *et al.* (2016a). Update: Interim Guidance for Preconception Counseling and Prevention of Sexual Transmission of Zika Virus for Persons with Possible Zika Virus Exposure - United States, September 2016. *MMWR Morb Mortal Wkly Rep* 65, 1077-1081.

Petersen, L.R., Jamieson, D.J., Powers, A.M., and Honein, M.A. (2016b). Zika Virus. *N Engl J Med* 374, 1552-1563.

Pethel, M., Falgout, B., and Lai, C.J. (1992). Mutational analysis of the octapeptide sequence motif at the NS1-NS2A cleavage junction of dengue type 4 virus. *J Virol* 66, 7225-7231.

Phoo, W.W., Li, Y., Zhang, Z., Lee, M.Y., Loh, Y.R., Tan, Y.B., Ng, E.Y., Lescar, J., Kang, C., and Luo, D. (2016). Structure of the NS2B-NS3 protease from Zika virus after self-cleavage. *Nat Commun* 7, 13410.

Pierce, K.K., Durbin, A.P., Grier, T., Tibery, C., Janiak, A., Lovchik, J., Jarvis, A., Carmolli, M., Kenney, H., Pletnev, A., *et al.* (2017). THE LIVE ATTENUATED CHIMERIC VACCINE RWN/DEN4 Delta 30 IS WELL-TOLERATED AND IMMUNOGENIC IN HEALTHY FLAVIVIRUS-NAIVE ADULT VOLUNTEERS 50-65 YEARS OF AGE. *Am J Trop Med Hyg* 95, 189-189.

Pinto, A.K., Richner, J.M., Poore, E.A., Patil, P.P., Amanna, I.J., Slifka, M.K., and Diamond, M.S. (2013). A Hydrogen Peroxide-Inactivated Virus Vaccine Elicits Humoral and Cellular Immunity and Protects against Lethal West Nile virus Infection in Aged Mice. *J Virol* 87, 1926-1936.

Pletnev, A.G., Swayne, D.E., Speicher, J., Rumentantsev, A.A., and Murphy, B.R. (2006). Chimeric West Nile/dengue virus vaccine candidate: Preclinical evaluation in mice, geese and monkeys for safety and immunogenicity. *Vaccine* 24, 6392-6404.

Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003). A graphical user interface to the CCP4 program suite. *Acta Crystallogr D Biol Crystallogr* 59, 1131-1137.

Powell, H.R., Kara, T., and Jones, C.L. (2007). Early experience with conversion to sirolimus in a pediatric renal transplant population. *Pediatr Nephrol* 22, 1773-1777.

Prusis, P., Junaid, M., Petrovska, R., Yahorava, S., Yahorau, A., Katzenmeier, G., Lapins, M., and Wikberg, J.E. (2013). Design and evaluation of substrate-based octapeptide and non substrate-based tetrapeptide inhibitors of dengue virus NS2B-NS3 proteases. *Biochem Biophys Res Commun* 434, 767-772.

Prusis, P., Lapins, M., Yahorava, S., Petrovska, R., Niyomrattanakit, P., Katzenmeier, G., and Wikberg, J.E.S. (2008). Proteochemometrics analysis of substrate interactions with dengue virus NS3 proteases. *Bioorg Med Chem* 16, 9369-9377.

PyMOL The PyMOL Molecular Graphics System, V., Schrödinger, LLC.

Rahuel, J., Rasetti, V., Maibaum, J., Rueger, H., Goschke, R., Cohen, N.C., Stutz, S., Cumin, F., Fuhrer, W., Wood, J.M., *et al.* (2000). Structure-based drug design: the discovery of novel nonpeptide orally active inhibitors of human renin. *Chem Biol* 7, 493-504.

Ramos, H.J., Lanteri, M.C., Blahnik, G., Negash, A., Suthar, M.S., Brassil, M.M., Sodhi, K., Treuting, P.M., Busch, M.P., Norris, P.J., *et al.* (2012). IL-1beta signaling promotes CNS-intrinsic immune control of West Nile virus infection. *PLoS Pathog* 8, e1003039.

Ramsauer, K., Schrauf, S., Vielnascher, R., Kort, A., Mullner, M., and Tauber, E. (2017). Phase 1 Study of Mv-Zika, a Live Recombinant Measles Virus Vaccine to Prevent Zika Virus Infection. *Am J Trop Med Hyg* 97, 45-45.

Randolph, V.B., Winkler, G., and Stollar, V. (1990). Acidotropic amines inhibit proteolytic processing of flavivirus prM protein. *Virology* 174, 450-458.

Rasmussen, S.A., Jamieson, D.J., Honein, M.A., and Petersen, L.R. (2016). Zika Virus and Birth Defects--Reviewing the Evidence for Causality. *N Engl J Med* 374, 1981-1987.

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

Rey, F.A., Heinz, F.X., Mandl, C., Kunz, C., and Harrison, S.C. (1995b). The envelope glycoprotein from tick-borne encephalitis virus at 2 A resolution. *Nature* 375, 291-298.

Rey, F.A., Stiasny, K., and Heinz, F.X. (2017). Flavivirus structural heterogeneity: implications for cell entry. *Current Opinion in Virology* 24, 132-139.

Rey, F.A., Stiasny, K., Vaney, M.C., Dellarole, M., and Heinz, F.X. (2018). The bright and the dark side of human antibody responses to flaviviruses: lessons for vaccine design. *EMBO reports* 19, 206-224.

Reyes-Del Valle, J., Chavez-Salinas, S., Medina, F., and Del Angel, R.M. (2005). Heat shock protein 90 and heat shock protein 70 are components of dengue virus receptor complex in human cells. *J Virol* 79, 4557-4567.

Richner, J.M., Himansu, S., Dowd, K.A., Butler, S.L., Salazar, V., Fox, J.M., Julander, J.G., Tang, W.W., Shrestha, S., Pierson, T.C., *et al.* (2017). Modified mRNA Vaccines Protect against Zika Virus Infection (vol 168, pg 1114, 2017). *Cell* 169, 176-176.

Robin, G., Chappell, K., Stoermer, M.J., Hu, S.H., Young, P.R., Fairlie, D.P., and Martin, J.L. (2009). Structure of West Nile virus NS3 protease: ligand stabilization of the catalytic conformation. *J Mol Biol* 385, 1568-1577.

Rodenhuis-Zybert, I.A., van der Schaar, H.M., da Silva Voorham, J.M., van der Ende-Metselaar, H., Lei, H.Y., Wilschut, J., and Smit, J.M. (2010). Immature dengue virus: a veiled pathogen? *PLoS Pathog* 6, e1000718.

Rodenhuis-Zybert, I.A., Wilschut, J., and Smit, J.M. (2011). Partial maturation: an immune-evasion strategy of dengue virus? *Trends in Microbiology* 19, 248-254.

Rothan, H.A., Mohamed, Z., Paydar, M., Rahman, N.A., and Yusof, R. (2014). Inhibitory effect of doxycycline against dengue virus replication in vitro. *Arch Virol* 159, 711-718.

Rothan, H.A., Mohamed, Z., Suhaeb, A.M., Rahman, N.A., and Yusof, R. (2013). Antiviral cationic peptides as a strategy for innovation in global health therapeutics for dengue virus: high yield production of the biologically active recombinant plectasin peptide. *OMICS* 17, 560-567.

Roundy, C.M., Azar, S.R., Brault, A.C., Ebel, G.D., Failloux, A.B., Fernandez-Salas, I., Kitron, U., Kramer, L.D., Lourenco-de-Oliveira, R., Osorio, J.E., *et al.* (2017). Lack of evidence for Zika virus transmission by Culex mosquitoes. *Emerg Microbes Infect* 6, e90.

Sakoonwatanyoo, P., Boonsanay, V., and Smith, D.R. (2006). Growth and production of the dengue virus in C6/36 cells and identification of a laminin-binding protein as a candidate serotype 3 and 4 receptor protein. *Intervirology* 49, 161-172.

Salazar, M.I., Richardson, J.H., Sanchez-Vargas, I., Olson, K.E., and Beaty, B.J. (2007). Dengue virus type 2: replication and tropisms in orally infected *Aedes aegypti* mosquitoes. *BMC Microbiol* 7, 9.

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

Samuel, M.A., and Diamond, M.S. (2005). Alpha/beta interferon protects against lethal West Nile virus infection by restricting cellular tropism and enhancing neuronal survival. *J Virol* 79, 13350-13361.

Sbardellati, A., Scarselli, E., Amati, V., Falcinelli, S., Kekule, A.S., and Traboni, C. (2000). Processing of GB virus B non-structural proteins in cultured cells requires both NS3 protease and NS4A cofactor. *J Gen Virol* 81, 2183-2188.

Schmidt, A.C., Lin, L., Martinez, L.J., Ruck, R.C., Eckels, K.H., Collard, A., De La Barrera, R., Paolino, K.M., Toussaint, J.F., Lepine, E., *et al.* (2017). Phase 1 Randomized Study of a Tetravalent Dengue Purified Inactivated Vaccine in Healthy Adults in the United States. *Am J Trop Med Hyg* 96, 1325-1337.

Schuessler, A., Funk, A., Lazear, H.M., Cooper, D.A., Torres, S., Daffis, S., Jha, B.K., Kumagai, Y., Takeuchi, O., Hertzog, P., *et al.* (2012). West Nile virus Noncoding Subgenomic RNA Contributes to Viral Evasion of the Type I Interferon-Mediated Antiviral Response. *J Virol* 86, 5708-5718.

Schuller, A., Yin, Z., Brian Chia, C.S., Doan, D.N., Kim, H.K., Shang, L., Loh, T.P., Hill, J., and Vasudevan, S.G. (2011). Tripeptide inhibitors of dengue and West Nile virus NS2B-NS3 protease. *Antiviral Res* 92, 96-101.

Schuttelkopf, A.W., and van Aalten, D.M.F. (2004). PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr* 60, 1355-1363.

Schweizer, M., and Peterhans, E. (2014). Pestiviruses. *Annu Rev Anim Biosci* 2, 141-+.

Shan, X.Y., Wang, X.M., Yuan, Q., Zheng, Y.P., Zhang, H.H., Wu, Y.H., and Yang, J. (2015). Evaluation of the diagnostic accuracy of nonstructural protein 1 Ag-based tests for dengue virus in Asian population: a meta-analysis. *Bmc Infectious Diseases* 15.

Shannon, A.E., Pedroso, M.M., Chappell, K.J., Watterson, D., Liebscher, S., Kok, W.M., Fairlie, D.P., Schenk, G., and Young, P.R. (2016). Product release is rate-limiting for catalytic processing by the Dengue virus protease. *Sci Rep* 6, 37539.

Shekarian, T., Valsesia-Wittmann, S., Brody, J., Michallet, M.C., Depil, S., Caux, C., and Marabelle, A. (2017). Pattern recognition receptors: immune targets to enhance cancer immunotherapy. *Ann Oncol* 28, 1756-1766.

Shi, Y.G. (2014). A Glimpse of Structural Biology through X-Ray Crystallography. *Cell* 159, 995-1014.

Shiryaev, S.A., Cheltsov, A.V., Gawlik, K., Ratnikov, B.I., and Strongin, A.Y. (2011). Virtual Ligand Screening of the National Cancer Institute (NCI) Compound Library Leads to the Allosteric Inhibitory Scaffolds of the West Nile virus NS3 Proteinase. *Assay and Drug Development Technologies* 9, 69-78.

Shiryaev, S.A., Chernov, A.V., Aleshin, A.E., Shiryaeva, T.N., and Strongin, A.Y. (2009). NS4A regulates the ATPase activity of the NS3 helicase: a novel cofactor role of the non-structural protein NS4A from West Nile virus. *J Gen Virol* 90, 2081-2085.

Shiryaev, S.A., Ratnikov, B.I., Aleshin, A.E., Kozlov, I.A., Nelson, N.A., Lebl, M., Smith, J.W., Liddington, R.C., and Strongin, A.Y. (2007). Switching the substrate specificity of the two-component NS2B-NS3 flavivirus proteinase by structure-based mutagenesis. *J Virol* 81, 4501-4509.

Shu, P.Y., and Huang, J.H. (2004). Current advances in dengue diagnosis. *Clinical and Diagnostic Laboratory Immunology* 11, 642-650.

Simmonds, P., Becher, P., Bukh, J., Gould, E.A., Meyers, G., Monath, T., Muerhoff, S., Pletnev, A., Rico-Hesse, R., Smith, D.B., *et al.* (2017). ICTV Virus Taxonomy Profile: Flaviviridae. *J Gen Virol* 98, 2-3.

Simons, J.N., Desai, S.M., Schultz, D.E., Lemon, S.M., and Mushahwar, I.K. (1996). Translation initiation in GB viruses A and C: evidence for internal ribosome entry and implications for genome organization. *J Virol* 70, 6126-6135.

Sirivichayakul, C., Barranco-Santana, E.A., Esquelin-Rivera, I., Oh, H.M.L., Raanan, M., Sariol, C.A., Shek, L.P., Simasathien, S., Smith, M.K., Velez, I.D., *et al.* (2016). Safety and Immunogenicity of a Tetravalent Dengue Vaccine Candidate in Healthy Children and Adults in Dengue-Endemic Regions: A Randomized, Placebo-Controlled Phase 2 Study. *J Infect Dis* 213, 1562-1572.

Sirohi, D., and Kuhn, R.J. (2017). Zika Virus Structure, Maturation, and Receptors. *J Infect Dis* 216, S935-S944.

Sohn, Y.M. (2000). Japanese encephalitis immunization in South Korea: past, present, and future. *Emerg Infect Dis* 6, 17-24.

Spahn, C.M.T., Kieft, J.S., Grassucci, R.A., Penczek, P.A., Zhou, K.H., Doudna, J.A., and Frank, J. (2001). Hepatitis C virus IRES RNA-induced changes in the conformation of the 40S ribosomal subunit. *Science* 291, 1959-1962.

Sriurairatna, S., and Bhamarapravati, N. (1977). Replication of dengue-2 virus in *Aedes albopictus* mosquitoes. An electron microscopic study. *Am J Trop Med Hyg* 26, 1199-1205.

Staessen, J.A., Li, Y., and Richart, T. (2006). Oral renin inhibitors. *Lancet* 368, 1449-1456.

Stapleton, J.T., Fong, S., Muerhoff, A.S., Bukh, J., and Simmonds, P. (2011). The GB viruses: a review and proposed classification of GBV-A, GBV-C (HGV), and GBV-D in genus Pegivirus within the family Flaviviridae. *J Gen Virol* 92, 233-246.

Stern, O., Hung, Y.F., Valda, O., Yaffe, Y., Harris, E., Hoffmann, S., Willbold, D., and Sklan, E.H. (2013). An N-terminal amphipathic helix in dengue virus nonstructural protein 4A mediates oligomerization and is essential for replication. *J Virol* 87, 4080-4085.

Steuer, C., Gege, C., Fischl, W., Heinonen, K.H., Bartenschlager, R., and Klein, C.D. (2011). Synthesis and biological evaluation of alpha-ketoamides as inhibitors of the Dengue virus protease with antiviral activity in cell-culture. *Bioorg & Med Chem* 19, 4067-4074.

Stevens, R.C. (2000). Design of high-throughput methods of protein production for structural biology. *Structure* 8, R177-185.

Stocks, C.E., and Lobigs, M. (1998). Signal peptidase cleavage at the flavivirus C-prM junction: dependence on the viral NS2B-3 protease for efficient processing requires determinants in C, the signal peptide, and prM. *J Virol* 72, 2141-2149.

Stokes, A., Bauer, J.H., Hudson, N.P. (1928). Experimental transmission of yellow fever to laboratory animals. *Am J Trop Med* 103-164, 8.

Su, X.C., Ozawa, K., Qi, R., Vasudevan, S.G., Lim, S.P., and Otting, G. (2009a). NMR analysis of the dynamic exchange of the NS2B cofactor between open and closed conformations of the West Nile virus NS2B-NS3 protease. *PLoS Negl Trop Dis* 3, e561.

Su, X.C., Ozawa, K., Yagi, H., Lim, S.P., Wen, D., Ekonomiuk, D., Huang, D., Keller, T.H., Sonntag, S., Caflisch, A., *et al.* (2009b). NMR study of complexes between low molecular mass inhibitors and the West Nile virus NS2B-NS3 protease. *FEBS J* 276, 4244-4255.

Szretter, K.J., Daniels, B.P., Cho, H., Gainey, M.D., Yokoyama, W.M., Gale, M., Jr., Virgin, H.W., Klein, R.S., Sen, G.C., and Diamond, M.S. (2012). 2'-O methylation of the viral mRNA cap by West Nile virus evades ifit1-dependent and -independent mechanisms of host restriction in vivo. *PLoS Path* 8, e1002698.

Tadano, M., Makino, Y., Fukunaga, T., Okuno, Y., and Fukai, K. (1989). Detection of Dengue-4 Virus Core Protein in the Nucleus .1. A Monoclonal-Antibody to Dengue-4 Virus Reacts with the Antigen in the Nucleus and Cytoplasm. *J Gen Virol* 70, 1409-1415.

Takikawa, S., Engle, R.E., Emerson, S.U., Purcell, R.H., St Claire, M., and Bukh, J. (2006). Functional analyses of GB virus B p13 protein: development of a recombinant GB virus B hepatitis virus with a p7 protein. *Proc Natl Acad Sci U S A* 103, 3345-3350.

Tassaneetrithep, B., Burgess, T.H., Granelli-Piperno, A., Trumpfheller, C., Finke, J., Sun, W., Eller, M.A., Pattanapanyasat, K., Sarasombath, S., Birk, D.L., *et al.* (2003). DC-SIGN (CD209) mediates dengue virus infection of human dendritic cells. *J Exp Med* 197, 823-829.

Tautz, N., Tews, B.A., and Meyers, G. (2015). The Molecular Biology of Pestiviruses. *Advances in Virus Research*, Vol 93 93, 47-160.

Tay, M.Y., Saw, W.G., Zhao, Y., Chan, K.W., Singh, D., Chong, Y., Forwood, J.K., Ooi, E.E., Gruber, G., Lescar, J., *et al.* (2015). The C-terminal 50 amino acid residues of dengue NS3 protein are important for NS3-NS5 interaction and viral replication. *J Biol Chem* 290, 2379-2394.

Tay, M.Y.F., Smith, K., Ng, I.H.W., Chan, K.W.K., Zhao, Y.Q., Ooi, E.E., Lescar, J., Luo, D.H., Jans, D.A., Forwood, J.K., *et al.* (2016). The C-terminal 18 Amino Acid Region of Dengue Virus NS5 Regulates its Subcellular Localization and Contains a Conserved Arginine Residue Essential for Infectious Virus Production. *PLoS Path* 12.

Theiler, M., and Smith, H.H. (1937). THE USE OF YELLOW FEVER VIRUS MODIFIED BY IN VITRO CULTIVATION FOR HUMAN IMMUNIZATION. *J Exper Med.* 65, 787-800.

Theodore, D., and Lemon, S.M. (1997). GB virus C, hepatitis G virus, or human orphan flavivirus? *Hepatology* 25, 1285-1286.

Thepparit, C., and Smith, D.R. (2004). Serotype-specific entry of dengue virus into liver cells: Identification of the 37-kilodalton/67-kilodalton high-affinity laminin receptor as a dengue virus serotype 1 receptor. *J Virol* 78, 12647-12656.

Tian, H., Ji, X., Yang, X., Xie, W., Yang, K., Chen, C., Wu, C., Chi, H., Mu, Z., Wang, Z., *et al.* (2016a). The crystal structure of Zika virus helicase: basis for antiviral drug design. *Protein Cell* 7, 450-454.

Tian, H., Ji, X., Yang, X., Zhang, Z., Lu, Z., Yang, K., Chen, C., Zhao, Q., Chi, H., Mu, Z., *et al.* (2016b). Structural basis of Zika virus helicase in recognizing its substrates. *Protein Cell* 7, 562-570.

Tickle, I., Sharff, A., Vinkovic, M., Yon, J., and Jhoti, H. (2004). High-throughput protein crystallography and drug discovery. *Chemical Society Reviews* 33, 558-565.

Tomlinson, S.M., Malmstrom, R.D., Russo, A., Mueller, N., Pang, Y.P., and Watowich, S.J. (2009). Structure-based discovery of dengue virus protease inhibitors. *Antiviral Res* 82, 110-114.

Tomlinson, S.M., and Watowich, S.J. (2012). Use of parallel validation high-throughput screens to reduce false positives and identify novel dengue NS2B-NS3 protease inhibitors. *Antiviral Res* 93, 245-252.

Tsai, Y.T., Chang, S.Y., Lee, C.N., and Kao, C.L. (2009). Human TLR3 recognizes dengue virus and modulates viral replication in vitro. *Cellular Microbiology* 11, 604-615.

Valle, R.P., and Falgout, B. (1998). Mutagenesis of the NS3 protease of dengue virus type 2. *J Virol* 72, 624-632.

van der Schaar, H.M., Rust, M.J., Waarts, B.L., van der Ende-Metselaar, H., Kuhn, R.J., Wilschut, J., Zhuang, X., and Smit, J.M. (2007). Characterization of the early events in dengue virus cell entry by biochemical assays and single-virus tracking. *J Virol* 81, 12019-12028.

Vasilakis, N., and Weaver, S.C. (2017). Flavivirus transmission focusing on Zika. *Curr Opin Virol* 22, 30-35.

Vesnaver, T.V., Tul, N., Mehrabi, S., Parissone, F., Strafela, P., Mlakar, J., Pizem, J., Korva, M., Zupanc, T.A., and Popovic, M. (2017). Zika virus associated microcephaly/micrencephaly-fetal brain imaging in comparison with neuropathology. *BJOG* 124, 521-525.

Villar, L., Dayan, G.H., Arredondo-Garcia, J.L., Rivera, D.M., Cunha, R., Deseda, C., Reynales, H., Costa, M.S., Morales-Ramirez, J.O., Carrasquilla, G., *et al.* (2015). Efficacy of a tetravalent dengue vaccine in children in Latin America. *N Engl J Med* 372, 113-123.

Vogel, G. (2016). One year later, Zika scientists prepare for a long war. *Science* 354, 1088-1089.

Wan, S.W., Lin, C.F., Wang, S., Chen, Y.H., Yeh, T.M., Liu, H.S., Anderson, R., and Lin, Y.S. (2013). Current progress in dengue vaccines. *J Biomed Sci* 20, 37.

Wang, L., Chen, R.F., Liu, J.W., Lee, I.K., Lee, C.P., Kuo, H.C., Huang, S.K., and Yang, K.D. (2011). DC-SIGN (CD209) Promoter -336 A/G polymorphism is associated with dengue hemorrhagic fever and correlated to DC-SIGN expression and immune augmentation. *PLoS neglected tropical diseases* 5, e934.

Weaver, S.C., and Barrett, A.D.T. (2004). Transmission cycles, host range, evolution and emergence of arboviral disease. *Nat Rev Microbiol* 2, 789-801.

Weaver, S.C., Costa, F., Garcia-Blanco, M.A., Ko, A.I., Ribeiro, G.S., Saade, G., Shi, P.Y., and Vasilakis, N. (2016). Zika virus: History, emergence, biology, and prospects for control. *Antiviral Res* 130, 69-80.

Weigel, L.F., Nitsche, C., Graf, D., Bartenschlager, R., and Klein, C.D. (2015). Phenylalanine and Phenylglycine Analogues as Arginine Mimetics in Dengue Protease Inhibitors. *J Med Chem* 58, 7719-7733.

Weinert, T., Olieric, V., Waltersperger, S., Panepucci, E., Chen, L., Zhang, H., Zhou, D., Rose, J., Ebihara, A., Kuramitsu, S., *et al.* (2015). Fast native-SAD phasing for routine macromolecular structure determination. *Nat Methods* 12, 131-133.

Welsch, S., Miller, S., Romero-Brey, I., Merz, A., Bleck, C.K.E., Walther, P., Fuller, S.D., Antony, C., Krijnse-Locker, J., and Bartenschlager, R. (2009). Composition and Three-Dimensional Architecture of the Dengue Virus Replication and Assembly Sites. *Cell Host & Microbe* 5, 365-375.

Wen, J., and Shresta, S. (2017). ADE-ing and Abetting Zika. *Cell Host Microbe* 21, 557-558.

Werme, K., Wigerius, M., and Johansson, M. (2008). Tick-borne encephalitis virus NS5 associates with membrane protein scribble and impairs interferon-stimulated JAK-STAT signalling. *Cell Microbiol* 10, 696-712.

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

Wilson, J.R., de Sessions, P.F., Leon, M.A., and Scholle, F. (2008). West Nile virus nonstructural protein 1 inhibits TLR3 signal transduction. *J Virol* 82, 8262-8271.

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.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A., *et al.* (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67, 235-242.

Wood, J.M., Maibaum, J., Rahuel, J., Grutter, M.G., Cohen, N.C., Rasetti, V., Ruger, H., Goschke, R., Stutz, S., Fuhrer, W., *et al.* (2003). Structure-based design of aliskiren, a novel orally effective renin inhibitor. *Biochem Biophys Res Commun* 308, 698-705.

Wu, J., Bera, A.K., Kuhn, R.J., and Smith, J.L. (2005). Structure of the Flavivirus helicase: implications for catalytic activity, protein interactions, and proteolytic processing. *J Virol* 79, 10268-10277.

Xiang, J., Klinzman, D., McLinden, J., Schmidt, W.N., LaBrecque, D.R., Gish, R., and Stapleton, J.T. (1998). Characterization of hepatitis G virus (GB-C virus) particles: evidence for a nucleocapsid and expression of sequences upstream of the E1 protein. *J Virol* 72, 2738-2744.

Xiang, J., McLinden, J.H., Chang, Q., Jordan, E.L., and Stapleton, J.T. (2008). Characterization of a peptide domain within the GB virus C NS5A phosphoprotein that inhibits HIV replication. *PLoS One* 3, e2580.

Xie, X., Zou, J., Puttikhunt, C., Yuan, Z., and Shi, P.Y. (2015). Two distinct sets of NS2A molecules are responsible for dengue virus RNA synthesis and virion assembly. *J Virol* 89, 1298-1313.

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

Xu, T., Sampath, A., Chao, A., Wen, D., Nanao, M., Luo, D., Chene, P., Vasudevan, S.G., and Lescar, J. (2006). Towards the design of flavivirus helicase/NTPase inhibitors: crystallographic and mutagenesis studies of the dengue virus NS3 helicase catalytic domain. *Novartis Found Symp* 277, 87-97; discussion 97-101, 251-103.

Yang, Y., Meng, Y., Halloran, M.E., and Longini, I.M. (2018). Dengvaxia Efficacy Dependency on Serostatus: A Closer Look at More Recent Data Reply. *Clinical Infectious Diseases* 66, 642-642.

Yao, N.H., Reichert, P., Taremi, S.S., Prosser, W.W., and Weber, P.C. (1999). Molecular views of viral polyprotein processing revealed by the crystal structure of the hepatitis C virus bifunctional protease-helicase. *Structure with Folding & Design* 7, 1353-1363.

Yap, L.J., Luo, D., Chung, K.Y., Lim, S.P., Bodenreider, C., Noble, C., Shi, P.Y., and Lescar, J. (2010). Crystal structure of the dengue virus methyltransferase bound to a 5'-capped octameric RNA. *PLoS One* 5.

Yildiz, M., Ghosh, S., Bell, J.A., Sherman, W., and Hardy, J.A. (2013). Allosteric inhibition of the NS2B-NS3 protease from dengue virus. *ACS Chem Biol* 8, 2744-2752.

Yin, Z., Patel, S.J., Wang, W.L., Chan, W.L., Ranga Rao, K.R., Wang, G., Ngew, X., Patel, V., Beer, D., Knox, J.E., *et al.* (2006a). Peptide inhibitors of dengue virus NS3 protease. Part 2: SAR study of tetrapeptide aldehyde inhibitors. *Bioorg Med Chem Lett* 16, 40-43.

Yin, Z., Patel, S.J., Wang, W.L., Wang, G., Chan, W.L., Rao, K.R., Alam, J., Jeyaraj, D.A., Ngew, X., Patel, V., *et al.* (2006b). Peptide inhibitors of Dengue virus NS3 protease. Part 1: Warhead. *Bioorg Med Chem Lett* 16, 36-39.

Yoo, B.J., Spaete, R.R., Geballe, A.P., Selby, M., Houghton, M., and Han, J.H. (1992). 5' end-dependent translation initiation of hepatitis C viral RNA and the presence of putative positive and negative translational control elements within the 5' untranslated region. *Virology* 191, 889-899.

Youn, S., Li, T., McCune, B.T., Edeling, M.A., Fremont, D.H., Cristea, I.M., and Diamond, M.S. (2012). Evidence for a genetic and physical interaction between nonstructural proteins NS1 and NS4B that modulates replication of West Nile virus. *J Virol* 86, 7360-7371.

Yu, I.M., Zhang, W., Holdaway, H.A., Li, L., Kostyuchenko, V.A., Chipman, P.R., Kuhn, R.J., Rossmann, M.G., and Chen, J. (2008). Structure of the immature dengue virus at low pH primes proteolytic maturation. *Science* 319, 1834-1837.

Yun, S.I., and Lee, Y.M. (2014). Japanese encephalitis: the virus and vaccines. *Hum Vaccin Immunother* 10, 263-279.

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

Zamani, M., and Zamani, V. (2017). Sexual Transmission of Zika Virus: An Assessment of the Evidence. *Iran J Public Health* 46, 1305-1306.

Zhang, C.Y., Feng, T.T., Cheng, J.B., Li, Y.J., Yin, X.Y., Zeng, W.H., Jin, X.Y., Li, Y.L., Guo, F., and Jin, T.C. (2017). Structure of the NS5 methyltransferase from Zika virus and implications in inhibitor design. *Biochem Biophys Res Comm* 492, 624-630.

Zhang, X., Sheng, J., Austin, S.K., Hoornweg, T.E., Smit, J.M., Kuhn, R.J., Diamond, M.S., and Rossmann, M.G. (2015). Structure of acidic pH dengue virus showing the fusogenic glycoprotein trimers. *J Virol* 89, 743-750.

Zhang, Y., Corver, J., Chipman, P.R., Zhang, W., Pletnev, S.V., Sedlak, D., Baker, T.S., Strauss, J.H., Kuhn, R.J., and Rossmann, M.G. (2003). Structures of immature flavivirus particles. *EMBO J* 22, 2604-2613.

Zhang, Y., Zhang, W., Ogata, S., Clements, D., Strauss, J.H., Baker, T.S., Kuhn, R.J., and Rossmann, M.G. (2004). Conformational changes of the flavivirus E glycoprotein. *Structure* 12, 1607-1618.

Zhang, Z., Li, Y., Loh, Y.R., Phoo, W.W., Hung, A.W., Kang, C., and Luo, D. (2016a). Crystal structure of unlinked NS2B-NS3 protease from Zika virus. *Science* 354, 1597-1600.

Zhang, Z.Z., Li, Y., Loh, Y.R., Phoo, W.W., Hung, A.W., Kang, C.B., and Luo, D.H. (2016b). Crystal structure of unlinked NS2B-NS3 protease from Zika virus. *Science* 354, 1597-1600.

Zhao, Y., Soh, T.S., Chan, K.W., Fung, S.S., Swaminathan, K., Lim, S.P., Shi, P.Y., Huber, T., Lescar, J., Luo, D., *et al.* (2015a). Flexibility of NS5 Methyltransferase-Polymerase Linker Region Is Essential for Dengue Virus Replication. *J Virol* 89, 10717-10721.

Zhao, Y., Soh, T.S., Zheng, J., Chan, K.W., Phoo, W.W., Lee, C.C., Tay, M.Y., Swaminathan, K., Cornvik, T.C., Lim, S.P., *et al.* (2015b). A crystal structure of the Dengue virus NS5 protein reveals a novel inter-domain interface essential for protein flexibility and virus replication. *PLoS Pathog* 11, e1004682.

Zheng, H.P., Handing, K.B., Zimmerman, M.D., Shabalin, I.G., Almo, S.C., and Minor, W. (2015). X-ray crystallography over the past decade for novel drug discovery - where are we heading next? *Expert Opinion on Drug Discovery* 10, 975-989.

Zheng, H.P., Hou, J., Zimmerman, M.D., Wlodawer, A., and Minor, W. (2014). The future of crystallography in drug discovery. *Expert Opinion on Drug Discovery* 9, 125-137.

Zhou, G.C., Weng, Z., Shao, X., Liu, F., Nie, X., Liu, J., Wang, D., Wang, C., and Guo, K. (2013). Discovery and SAR studies of methionine-proline anilides as dengue virus NS2B-NS3 protease inhibitors. *Bioorg Med Chem Lett* 23, 6549-6554.

Zou, J., Xie, X., Lee le, T., Chandrasekaran, R., Reynaud, A., Yap, L., Wang, Q.Y., Dong, H., Kang, C., Yuan, Z., *et al.* (2014). Dimerization of flavivirus NS4B protein. *J Virol* 88, 3379-3391.

Full wwPDB X-ray Structure Validation Report ⓘ

Jul 29, 2016 – 04:15 PM JST

PDB ID : 5GJ4
Title : Structure of NS2B-NS3 Protease from Zika Virus caught after self-cleavage
Authors : Phoo, W.W.; Li, Y.
Deposited on : 2016-06-27
Resolution : 1.84 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>
with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : unknown
Xtriage (Phenix) : 1.9-1692
EDS : rb-20027939
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20027939

1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.84 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	2634 (1.86-1.82)
Clashscore	102246	2862 (1.86-1.82)
Ramachandran outliers	100387	2831 (1.86-1.82)
Sidechain outliers	100360	2832 (1.86-1.82)
RSRZ outliers	91569	2639 (1.86-1.82)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	61	<div> <div>61%</div> <div>10%</div> <div>30%</div> </div>
1	C	61	<div> <div>61%</div> <div>8%</div> <div>30%</div> </div>
1	E	61	<div> <div>51%</div> <div>18%</div> <div>30%</div> </div>
1	G	61	<div> <div>56%</div> <div>15%</div> <div>30%</div> </div>
2	B	177	<div> <div>73%</div> <div>14%</div> <div>14%</div> </div>
2	D	177	<div> <div>71%</div> <div>15%</div> <div>14%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	177	72% 14% 14%
2	H	177	71% 15% 14%

CONFIDENTIAL

VALIDATION REPORT

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 7043 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Serine protease subunit NS2B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	43	Total	C	N	O	S	0	0	0
			334	205	56	72	1			
1	C	43	Total	C	N	O	S	0	0	0
			334	205	56	72	1			
1	E	43	Total	C	N	O	S	0	0	0
			334	205	56	72	1			
1	G	43	Total	C	N	O	S	0	0	0
			334	205	56	72	1			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	41	GLY	-	expression tag	UNP A0A142IX72
A	42	SER	-	expression tag	UNP A0A142IX72
A	43	HIS	-	expression tag	UNP A0A142IX72
A	44	MET	-	expression tag	UNP A0A142IX72
C	41	GLY	-	expression tag	UNP A0A142IX72
C	42	SER	-	expression tag	UNP A0A142IX72
C	43	HIS	-	expression tag	UNP A0A142IX72
C	44	MET	-	expression tag	UNP A0A142IX72
E	41	GLY	-	expression tag	UNP A0A142IX72
E	42	SER	-	expression tag	UNP A0A142IX72
E	43	HIS	-	expression tag	UNP A0A142IX72
E	44	MET	-	expression tag	UNP A0A142IX72
G	41	GLY	-	expression tag	UNP A0A142IX72
G	42	SER	-	expression tag	UNP A0A142IX72
G	43	HIS	-	expression tag	UNP A0A142IX72
G	44	MET	-	expression tag	UNP A0A142IX72

- Molecule 2 is a protein called Serine protease NS3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	153	Total	C	N	O	S	0	0	0
			1144	724	202	213	5			
2	D	152	Total	C	N	O	S	0	0	0
			1139	721	201	212	5			
2	F	152	Total	C	N	O	S	0	1	0
			1140	721	200	214	5			
2	H	153	Total	C	N	O	S	0	0	0
			1144	724	202	213	5			

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	1	Total	Cl	0	0
			1	1		
3	D	1	Total	Cl	0	0
			1	1		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	89	Total	O	0	0
			89	89		
4	B	199	Total	O	0	0
			199	199		
4	C	81	Total	O	0	0
			81	81		
4	D	199	Total	O	0	0
			199	199		
4	E	101	Total	O	0	0
			101	101		
4	F	210	Total	O	0	0
			210	210		
4	G	75	Total	O	0	0
			75	75		
4	H	184	Total	O	0	0
			184	184		

3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ($RSRZ > 2$). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: Serine protease subunit NS2B

Chain A:

- Molecule 1: Serine protease subunit NS2B

Chain C:

- Molecule 1: Serine protease subunit NS2B

Chain E:

- Molecule 1: Serine protease subunit NS2B

Chain G:

- Molecule 2: Serine protease NS3

Chain B:

- Molecule 2: Serine protease NS3

Chain D:

71%

15%

•

14%

• Molecule 2: Serine protease NS3

Chain F:

72%

14%

•

14%

• Molecule 2: Serine protease NS3

Chain H:

71%

15%

•

14%

4 Data and refinement statistics

Property	Value	Source
Space group	P 31	Depositor
Cell constants a, b, c, α , β , γ	88.79 Å 88.79 Å 138.10 Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	31.50 – 1.84 31.50 – 1.84	Depositor EDS
% Data completeness (in resolution range)	97.8 (31.50-1.84) 94.6 (31.50-1.84)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.70 (at 1.84 Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, R_{free}	0.169 , 0.195 0.167 , 0.193	Depositor DCC
R_{free} test set	4846 reflections (4.84%)	DCC
Wilson B-factor (Å ²)	21.1	Xtriage
Anisotropy	0.012	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.28 , 32.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.39$, $\langle L^2 \rangle = 0.22$	Xtriage
Estimated twinning fraction	0.426 for -h,-k,l 0.487 for h,-h-k,-l 0.427 for -k,-h,-l	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	7043	wwPDB-VP
Average B, all atoms (Å ²)	29.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.10% of the height of the origin peak. No significant pseudotranslation is detected.*

¹ Intensities estimated from amplitudes.

² Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: CL

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z > 5	RMSZ	# Z > 5
1	A	0.30	0/337	0.53	0/452
1	C	0.30	0/337	0.51	0/452
1	E	0.30	0/337	0.52	0/452
1	G	0.30	0/337	0.52	0/452
2	B	0.31	0/1168	0.54	0/1585
2	D	0.31	0/1163	0.56	1/1578 (0.1%)
2	F	0.30	0/1167	0.54	0/1585
2	H	0.30	0/1168	0.54	0/1585
All	All	0.31	0/6014	0.54	1/8141 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	140	LEU	CA-CB-CG	5.33	127.55	115.30

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	334	0	315	9	0
1	C	334	0	315	6	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	E	334	0	315	15	0
1	G	334	0	315	16	0
2	B	1144	0	1139	21	0
2	D	1139	0	1137	27	0
2	F	1140	0	1130	23	0
2	H	1144	0	1139	30	0
3	B	1	0	0	0	0
3	D	1	0	0	0	0
4	A	89	0	0	6	0
4	B	199	0	0	12	3
4	C	81	0	0	4	0
4	D	199	0	0	9	3
4	E	101	0	0	8	1
4	F	210	0	0	8	2
4	G	75	0	0	9	0
4	H	184	0	0	10	3
All	All	7043	0	5805	124	6

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 11.

All (124) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:F:86:ASP:OD1	4:F:201:HOH:O	1.85	0.94
1:A:49:VAL:O	4:A:201:HOH:O	1.96	0.82
2:B:107:LYS:NZ	4:B:2105:HOH:O	2.12	0.81
1:A:49:VAL:N	4:A:202:HOH:O	2.13	0.80
2:B:167:GLN:O	4:B:2101:HOH:O	1.98	0.80
2:H:84:LYS:O	4:H:201:HOH:O	2.00	0.78
2:B:62:GLU:OE2	4:B:2102:HOH:O	2.02	0.78
1:G:49:VAL:HG12	4:G:227:HOH:O	1.84	0.77
2:F:66:ASP:OD2	4:F:202:HOH:O	2.01	0.77
2:H:143:CYS:HB2	2:H:145:ARG:HE	1.48	0.77
1:A:49:VAL:N	4:A:203:HOH:O	2.17	0.77
2:B:86:ASP:OD2	4:B:2103:HOH:O	2.02	0.76
2:D:108:ASN:HD21	2:H:105:ARG:HH12	1.33	0.76
2:B:118:THR:HG22	2:B:120:ASP:H	1.50	0.76
1:G:50:ASP:HA	4:G:205:HOH:O	1.84	0.76
2:B:84:LYS:NZ	4:B:2107:HOH:O	2.17	0.75
2:D:28:ARG:NH2	2:H:32:GLY:O	2.21	0.74

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:D:62:GLU:OE1	4:D:301:HOH:O	2.06	0.72
2:H:143:CYS:O	4:H:202:HOH:O	2.07	0.72
1:E:83:ASP:OD2	4:E:201:HOH:O	2.06	0.71
2:F:24:ARG:HH11	2:F:35:GLN:HE22	1.35	0.71
1:G:67:VAL:O	4:G:201:HOH:O	2.07	0.70
2:H:20:ASP:OD1	4:H:203:HOH:O	2.08	0.70
2:B:122:ASP:OD1	4:B:2104:HOH:O	2.09	0.69
1:E:49:VAL:HB	4:E:213:HOH:O	1.92	0.69
1:C:64:ASP:HB2	4:C:214:HOH:O	1.92	0.68
2:D:84:LYS:O	2:D:170:ARG:NH1	2.27	0.68
2:B:120:ASP:O	4:B:2106:HOH:O	2.13	0.65
1:E:49:VAL:HG21	4:F:365:HOH:O	1.98	0.64
1:E:72:PRO:HB3	4:E:244:HOH:O	2.00	0.61
2:F:42:GLN:NE2	4:F:213:HOH:O	2.33	0.61
2:D:108:ASN:HD22	2:H:105:ARG:HH22	1.49	0.60
2:D:64:ARG:NH2	4:D:303:HOH:O	2.11	0.60
2:B:157:LYS:O	4:B:2108:HOH:O	2.17	0.58
2:D:112:LEU:N	4:D:313:HOH:O	2.38	0.57
2:H:29:ARG:NE	4:H:208:HOH:O	2.26	0.57
2:B:129:ASP:O	4:B:2109:HOH:O	2.17	0.57
2:H:43:GLU:OE2	4:H:204:HOH:O	2.17	0.56
2:B:68:TYR:OH	4:B:2110:HOH:O	2.17	0.56
1:A:62:GLU:O	2:B:96:GLN:NE2	2.35	0.56
2:D:108:ASN:ND2	2:H:105:ARG:HH12	2.02	0.56
2:H:171:GLU:O	4:H:205:HOH:O	2.18	0.55
2:H:144:GLY:O	4:H:206:HOH:O	2.18	0.55
2:H:74:GLN:HG3	4:H:335:HOH:O	2.06	0.55
1:E:50:ASP:N	1:E:50:ASP:OD1	2.40	0.54
1:E:52:TYR:HB3	2:F:59:ARG:HB2	1.88	0.54
2:F:19:THR:N	4:F:219:HOH:O	2.40	0.54
1:A:86:LEU:HB2	4:A:228:HOH:O	2.07	0.53
1:G:130:ARG:OXT	4:G:202:HOH:O	2.19	0.53
1:E:65:ALA:O	4:E:202:HOH:O	2.18	0.53
1:C:60:THR:HG22	4:C:217:HOH:O	2.10	0.52
2:D:107:LYS:HD3	4:D:341:HOH:O	2.08	0.52
1:E:76:VAL:HG12	1:E:86:LEU:HD12	1.92	0.52
1:C:130:ARG:C	2:D:135:SER:HB2	2.30	0.52
1:G:63:LYS:NZ	4:G:211:HOH:O	2.42	0.51
2:D:28:ARG:NH1	4:D:304:HOH:O	2.11	0.51
2:D:145:ARG:NH2	4:D:319:HOH:O	2.43	0.51
1:A:63:LYS:NZ	4:A:209:HOH:O	2.42	0.51

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:E:127:THR:N	4:E:207:HOH:O	2.44	0.51
1:E:63:LYS:NZ	4:E:204:HOH:O	2.37	0.51
1:G:52:TYR:HB3	2:H:59:ARG:HB2	1.92	0.51
2:D:49:MET:HG3	2:D:135:SER:OG	2.10	0.50
1:C:64:ASP:O	4:C:201:HOH:O	2.18	0.50
2:D:118:THR:HG22	2:D:120:ASP:H	1.77	0.49
2:D:40:VAL:HG21	2:D:85:LEU:HD11	1.95	0.49
2:F:46:PHE:HB3	2:F:79:TYR:HB2	1.95	0.48
2:B:74:GLN:OE1	4:B:2111:HOH:O	2.19	0.48
1:G:130:ARG:O	2:H:133:GLY:N	2.43	0.48
1:G:69:GLY:HA2	4:G:246:HOH:O	2.13	0.48
1:G:75:ASP:OD1	4:G:203:HOH:O	2.20	0.48
2:D:90:ASP:N	4:D:309:HOH:O	2.31	0.48
2:D:46:PHE:HB3	2:D:79:TYR:HB2	1.96	0.47
1:E:130:ARG:C	2:F:135:SER:HB2	2.34	0.47
2:B:42:GLN:HG2	2:B:43:GLU:HG3	1.97	0.47
2:B:74:GLN:NE2	4:B:2101:HOH:O	2.17	0.47
2:F:39:GLY:HA3	2:F:46:PHE:CZ	2.49	0.47
1:E:130:ARG:HD2	2:F:130:TYR:O	2.16	0.46
2:H:19:THR:N	4:H:226:HOH:O	2.47	0.46
2:H:46:PHE:HB3	2:H:79:TYR:HB2	1.98	0.46
2:F:49:MET:HB2	2:F:52:VAL:HG23	1.99	0.45
2:D:28:ARG:HH22	2:H:29:ARG:H	1.63	0.45
1:A:67:VAL:HA	2:B:110:GLN:O	2.17	0.45
2:H:39:GLY:HA3	2:H:46:PHE:CZ	2.52	0.44
2:B:46:PHE:HB3	2:B:79:TYR:HB2	2.00	0.44
2:D:19:THR:N	4:D:325:HOH:O	2.50	0.44
2:D:40:VAL:HG11	2:D:138:PRO:HB3	2.00	0.44
2:F:66:ASP:OD1	4:F:205:HOH:O	2.21	0.44
2:B:39:GLY:HA3	2:B:46:PHE:CZ	2.53	0.43
2:D:50:TRP:NE1	4:D:305:HOH:O	2.22	0.43
2:F:51:HIS:O	4:F:203:HOH:O	2.21	0.43
2:H:143:CYS:SG	2:H:145:ARG:NH2	2.91	0.43
2:D:38:VAL:O	2:D:136:GLY:HA3	2.19	0.43
2:H:142:LYS:HG3	4:H:312:HOH:O	2.18	0.43
1:C:65:ALA:O	4:C:204:HOH:O	2.20	0.43
2:B:51:HIS:O	2:B:54:LYS:NZ	2.52	0.43
1:C:67:VAL:HA	2:D:110:GLN:O	2.19	0.43
1:A:127:THR:N	4:A:213:HOH:O	2.52	0.43
2:B:51:HIS:ND1	2:B:75:ASP:OD2	2.45	0.42
2:F:156:ILE:HG22	2:F:157:LYS:H	1.84	0.42

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:F:71[B]:ASP:HB3	2:F:74:GLN:HB2	2.01	0.42
2:H:38:VAL:O	2:H:136:GLY:HA3	2.20	0.42
1:E:49:VAL:HG22	4:E:273:HOH:O	2.18	0.42
1:E:52:TYR:CZ	2:F:26:MET:HB2	2.54	0.42
2:D:28:ARG:HH22	2:H:29:ARG:N	2.17	0.42
2:F:42:GLN:HG2	2:F:43:GLU:HG3	2.01	0.42
1:G:50:ASP:HB3	2:H:28:ARG:HB2	2.02	0.42
2:D:39:GLY:HA3	2:D:46:PHE:CZ	2.55	0.42
2:F:49:MET:HG3	2:F:135:SER:OG	2.20	0.42
2:F:156:ILE:HG22	2:F:157:LYS:N	2.35	0.42
1:G:49:VAL:N	4:G:218:HOH:O	2.52	0.42
1:G:130:ARG:HD2	2:H:130:TYR:O	2.20	0.41
2:D:28:ARG:NH2	2:H:29:ARG:O	2.54	0.41
2:F:165:ILE:O	4:F:206:HOH:O	2.21	0.41
2:H:82:PRO:HG2	2:H:84:LYS:HE2	2.02	0.41
2:D:89:TRP:CD1	2:D:90:ASP:N	2.89	0.41
1:G:49:VAL:O	4:G:205:HOH:O	2.22	0.41
2:F:84:LYS:HA	2:F:170:ARG:NH2	2.36	0.41
1:E:127:THR:O	4:E:203:HOH:O	2.22	0.40
2:F:84:LYS:HD3	2:F:170:ARG:HH22	1.86	0.40
1:G:52:TYR:CB	2:H:59:ARG:HB2	2.52	0.40
1:G:67:VAL:HA	2:H:110:GLN:O	2.22	0.40
1:G:52:TYR:CZ	2:H:26:MET:HB2	2.56	0.40
1:A:67:VAL:HG22	2:B:110:GLN:HB3	2.03	0.40
2:F:23:TYR:HB2	2:F:46:PHE:HE1	1.87	0.40

All (6) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
4:B:2204:HOH:O	4:F:366:HOH:O[3_554]	1.80	0.40
4:B:2204:HOH:O	4:F:344:HOH:O[3_554]	1.88	0.32
4:D:336:HOH:O	4:H:251:HOH:O[2_665]	2.09	0.11
4:B:2242:HOH:O	4:H:370:HOH:O[3_564]	2.12	0.08
4:D:485:HOH:O	4:H:384:HOH:O[2_665]	2.15	0.05
4:D:482:HOH:O	4:E:292:HOH:O[2_555]	2.18	0.02

5.3 Torsion angles [i](#)

5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	39/61 (64%)	39 (100%)	0	0	100	100
1	C	39/61 (64%)	38 (97%)	1 (3%)	0	100	100
1	E	39/61 (64%)	38 (97%)	1 (3%)	0	100	100
1	G	39/61 (64%)	39 (100%)	0	0	100	100
2	B	151/177 (85%)	147 (97%)	4 (3%)	0	100	100
2	D	150/177 (85%)	146 (97%)	4 (3%)	0	100	100
2	F	151/177 (85%)	145 (96%)	6 (4%)	0	100	100
2	H	151/177 (85%)	146 (97%)	5 (3%)	0	100	100
All	All	759/952 (80%)	738 (97%)	21 (3%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	36/51 (71%)	36 (100%)	0	100	100
1	C	36/51 (71%)	34 (94%)	2 (6%)	26	8
1	E	36/51 (71%)	34 (94%)	2 (6%)	26	8
1	G	36/51 (71%)	35 (97%)	1 (3%)	51	32
2	B	118/140 (84%)	116 (98%)	2 (2%)	68	54
2	D	118/140 (84%)	117 (99%)	1 (1%)	86	81

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	F	118/140 (84%)	117 (99%)	1 (1%)	86	81
2	H	118/140 (84%)	116 (98%)	2 (2%)	68	54
All	All	616/764 (81%)	605 (98%)	11 (2%)	66	52

All (11) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	140	LEU
2	B	158	ASN
1	C	51	MET
1	C	60	THR
2	D	140	LEU
1	E	50	ASP
1	E	68	THR
2	F	170	ARG
1	G	51	MET
2	H	119	LYS
2	H	120	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
2	B	158	ASN
2	D	108	ASN
2	F	35	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

6 Fit of model and data ⓘ

6.1 Protein, DNA and RNA chains ⓘ

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å ²)	Q<0.9
1	A	43/61 (70%)	-0.51	0 100 100	17, 32, 40, 45	0
1	C	43/61 (70%)	-0.54	0 100 100	18, 33, 42, 47	0
1	E	43/61 (70%)	-0.42	0 100 100	17, 32, 45, 53	0
1	G	43/61 (70%)	-0.42	0 100 100	18, 33, 43, 52	0
2	B	153/177 (86%)	-0.61	0 100 100	16, 25, 40, 59	0
2	D	152/177 (85%)	-0.62	0 100 100	16, 25, 36, 42	0
2	F	152/177 (85%)	-0.63	0 100 100	15, 25, 38, 51	0
2	H	153/177 (86%)	-0.64	0 100 100	17, 24, 37, 58	0
All	All	782/952 (82%)	-0.59	0 100 100	15, 26, 41, 59	0

There are no RSRZ outliers to report.

6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates ⓘ

There are no carbohydrates in this entry.

6.4 Ligands ⓘ

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors

of atoms in the group. The column labelled 'Q < 0.9' lists the number of atoms with occupancy less than 0.9.

CONFIDENTIAL VALIDATION REPORT

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors($\text{\AA}^2$)	Q<0.9
3	CL	D	201	1/1	0.99	0.07	-0.14	19,19,19,19	0
3	CL	B	2001	1/1	0.99	0.07	-0.64	19,19,19,19	0

6.5 Other polymers [i](#)

There are no such residues in this entry.

Full wwPDB X-ray Structure Validation Report ⓘ

Oct 14, 2016 – 10:14 AM JST

PDB ID : 5GPI
Title : Crystal Structures of Unlinked NS2B-NS3 Protease from Zika Virus and Its Complex with a Reverse Peptide Inhibitor
Deposited on : 2016-08-02
Resolution : 1.58 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report.

This report is produced by the wwPDB biocuration pipeline after annotation of the structure.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027939
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027939

1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.58 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	3815 (1.60-1.56)
Clashscore	102246	4131 (1.60-1.56)
Ramachandran outliers	100387	4021 (1.60-1.56)
Sidechain outliers	100360	4018 (1.60-1.56)
RSRZ outliers	91569	3823 (1.60-1.56)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	B	178	<div> <div>2%</div> <div> <div></div> <div>70%</div> <div>16%</div> <div>•</div> <div>13%</div> </div> </div>
1	D	178	<div> <div>%</div> <div> <div></div> <div>72%</div> <div>10%</div> <div>•</div> <div>17%</div> </div> </div>
1	F	178	<div> <div>4%</div> <div> <div></div> <div>66%</div> <div>19%</div> <div>•</div> <div>15%</div> </div> </div>
1	H	178	<div> <div>2%</div> <div> <div></div> <div>70%</div> <div>16%</div> <div>•</div> <div>14%</div> </div> </div>
2	A	53	<div> <div></div> <div> <div></div> <div>60%</div> <div>9%</div> <div>•</div> <div>28%</div> </div> </div>
2	C	53	<div> <div></div> <div> <div></div> <div>58%</div> <div>9%</div> <div>•</div> <div>30%</div> </div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	E	53	<div><div></div><div>4%</div><div>60%</div><div>11%</div><div>28%</div></div>
2	G	53	<div><div></div><div>68%</div><div>8%</div><div>25%</div></div>

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 6293 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called NS3 protease.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	H	153	Total	C	N	O	S	0	1	0
			1147	728	200	214	5			
1	D	148	Total	C	N	O	S	0	1	0
			1098	696	189	208	5			
1	B	154	Total	C	N	O	S	0	4	0
			1160	738	200	217	5			
1	F	152	Total	C	N	O	S	0	0	0
			1113	709	195	204	5			

- Molecule 2 is a protein called NS2B cofactor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	37	Total	C	N	O	S	0	0	0
			280	175	46	58	1			
2	A	38	Total	C	N	O	S	0	0	0
			280	174	46	59	1			
2	E	38	Total	C	N	O	S	0	0	0
			282	177	47	57	1			
2	G	40	Total	C	N	O	S	0	0	0
			299	185	48	65	1			

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	H	156	Total	O	0	0
			156	156		
3	C	27	Total	O	0	0
			27	27		
3	A	41	Total	O	0	0
			41	41		
3	E	33	Total	O	0	0
			33	33		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	G	52	Total 52	O 52	0	0
3	D	102	Total 102	O 102	0	0
3	B	127	Total 127	O 127	0	0
3	F	96	Total 96	O 96	0	0

3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ($RSRZ > 2$). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: NS3 protease

- Molecule 1: NS3 protease

- Molecule 1: NS3 protease

- Molecule 1: NS3 protease

- Molecule 2: NS2B cofactor

- Molecule 2: NS2B cofactor

- Molecule 2: NS2B cofactor

- Molecule 2: NS2B cofactor

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	60.27Å 60.38Å 215.46Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.17 – 1.58 46.17 – 1.58	Depositor EDS
% Data completeness (in resolution range)	99.5 (46.17-1.58) 97.9 (46.17-1.58)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.90 (at 1.58Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, R_{free}	0.166 , 0.196 0.166 , 0.196	Depositor DCC
R_{free} test set	5614 reflections (5.17%)	DCC
Wilson B-factor (Å ²)	18.3	Xtriage
Anisotropy	0.228	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 35.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.38$, $\langle L^2 \rangle = 0.20$	Xtriage
Estimated twinning fraction	0.326 for k,h,-l	Xtriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	6293	wwPDB-VP
Average B, all atoms (Å ²)	25.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 11.88% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# $ Z > 5$	RMSZ	# $ Z > 5$
1	B	0.28	0/1190	0.54	0/1619
1	D	0.27	0/1121	0.48	0/1523
1	F	0.27	0/1137	0.50	0/1546
1	H	0.27	0/1170	0.54	1/1585 (0.1%)
2	A	0.25	0/284	0.47	0/387
2	C	0.34	0/284	0.52	0/385
2	E	0.24	0/286	0.49	0/388
2	G	0.24	0/303	0.50	0/412
All	All	0.27	0/5775	0.51	1/7845 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	H	105	ARG	NE-CZ-NH1	-5.33	117.64	120.30

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	1160	0	1149	31	0
1	D	1098	0	1076	16	0
1	F	1113	0	1103	27	0
1	H	1147	0	1143	24	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	280	0	250	8	0
2	C	280	0	261	5	0
2	E	282	0	261	6	0
2	G	299	0	264	5	0
3	A	41	0	0	2	2
3	B	127	0	0	14	1
3	C	27	0	0	1	0
3	D	102	0	0	6	1
3	E	33	0	0	2	0
3	F	96	0	0	5	1
3	G	52	0	0	3	2
3	H	156	0	0	6	4
All	All	6293	0	5507	109	7

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 10.

All (109) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:D:106:ALA:O	3:D:201:HOH:O	1.88	0.90
1:B:129:ASP:OD2	3:B:201:HOH:O	1.89	0.89
1:B:17:GLU:OE1	3:B:202:HOH:O	1.98	0.81
1:H:104:GLU:OE2	3:H:201:HOH:O	1.99	0.79
1:B:31:LEU:CD2	1:B:32:GLY:H	1.97	0.78
1:B:59[B]:ARG:HH21	1:B:64:ARG:HB2	1.48	0.77
1:H:143:CYS:SG	3:H:315:HOH:O	2.43	0.77
1:H:24:ARG:HH11	1:H:35:GLN:HE22	1.31	0.77
1:B:145:ARG:NH2	3:B:204:HOH:O	2.19	0.75
1:B:123[B]:ILE:HD11	1:B:164:ALA:HB1	1.69	0.74
2:C:73:ARG:NH1	3:C:101:HOH:O	2.21	0.74
1:D:120:ASP:OD2	3:D:202:HOH:O	2.05	0.73
1:B:59[A]:ARG:NH1	3:B:206:HOH:O	2.21	0.73
1:B:17:GLU:N	3:B:208:HOH:O	2.23	0.71
1:B:24:ARG:HH11	1:B:35:GLN:HE22	1.39	0.69
1:B:31:LEU:HD23	1:B:32:GLY:H	1.58	0.69
1:D:156:ILE:HB	3:D:222:HOH:O	1.94	0.68
2:G:58:ASP:OD1	3:G:101:HOH:O	2.11	0.68
1:D:24:ARG:HH11	1:D:35:GLN:HE22	1.41	0.68
1:F:113:PRO:O	3:F:201:HOH:O	2.13	0.66
1:B:31:LEU:O	3:B:203:HOH:O	2.14	0.66

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:G:55:ARG:NH1	3:G:101:HOH:O	2.20	0.65
1:H:107:LYS:HD2	1:H:109:ILE:HD11	1.80	0.63
1:F:58:LEU:HB2	1:F:65:LEU:HB2	1.81	0.62
1:D:84:LYS:O	1:D:170:ARG:NH1	2.32	0.62
1:F:114:GLY:N	1:F:125:ALA:O	2.32	0.61
2:E:67:VAL:HG22	1:F:110:GLN:HB3	1.83	0.59
1:H:34:THR:N	3:H:207:HOH:O	2.36	0.59
1:B:17:GLU:N	3:B:213:HOH:O	2.35	0.58
1:H:96:GLN:HE21	1:H:110:GLN:HE21	1.49	0.58
1:D:114:GLY:N	1:D:125:ALA:O	2.36	0.57
1:H:59:ARG:NH1	3:H:205:HOH:O	2.35	0.57
1:F:29:ARG:O	3:F:202:HOH:O	2.18	0.57
2:E:81:SER:O	3:E:101:HOH:O	2.18	0.56
2:A:67:VAL:HG22	1:B:110:GLN:HB3	1.86	0.56
1:F:89:TRP:CZ2	1:F:113:PRO:HG2	2.41	0.56
2:C:55:ARG:NH2	2:C:58:ASP:OD1	2.31	0.56
1:F:89:TRP:CE2	1:F:113:PRO:HG2	2.41	0.56
1:F:24:ARG:HH11	1:F:35:GLN:HE22	1.54	0.55
1:F:156:ILE:HG22	1:F:157:LYS:H	1.71	0.55
1:F:64:ARG:NH1	1:F:66:ASP:OD1	2.38	0.55
1:B:160:SER:HB2	3:B:222:HOH:O	2.06	0.55
1:B:31:LEU:CD2	1:B:32:GLY:N	2.68	0.54
2:A:70:ASN:H	2:A:70:ASN:HD22	1.56	0.54
1:H:117:LYS:HD3	3:G:139:HOH:O	2.06	0.53
1:D:105:ARG:HD3	3:D:201:HOH:O	2.07	0.53
2:E:70:ASN:O	3:E:102:HOH:O	2.18	0.53
1:F:19:THR:HA	3:F:264:HOH:O	2.09	0.53
1:H:27:THR:HG22	1:H:34:THR:HB	1.90	0.53
1:B:98:LEU:HB3	1:B:138:PRO:HG2	1.91	0.52
2:C:51:MET:HG2	1:D:27:THR:HB	1.92	0.51
1:F:116:PHE:HB2	1:F:123:ILE:HG13	1.92	0.51
1:H:39:GLY:HA3	1:H:46:PHE:CZ	2.46	0.51
1:B:96:GLN:HE21	1:B:110:GLN:NE2	2.09	0.51
1:F:34:THR:HA	3:F:225:HOH:O	2.10	0.51
1:B:34:THR:HA	3:B:219:HOH:O	2.11	0.49
1:D:158:ASN:HD22	1:F:158:ASN:HD21	1.59	0.49
1:D:39:GLY:HA3	1:D:46:PHE:CZ	2.48	0.49
1:F:109:ILE:HG13	1:F:130:TYR:OH	2.13	0.48
1:H:116:PHE:HB2	1:H:123[A]:ILE:HG23	1.96	0.48
2:C:73:ARG:HH11	2:C:73:ARG:HG3	1.78	0.48
1:F:119:LYS:HG3	1:F:120:ASP:OD1	2.14	0.48

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:F:39:GLY:HA3	1:F:46:PHE:CZ	2.49	0.48
1:B:30:LEU:N	3:B:215:HOH:O	2.43	0.48
1:D:35:GLN:HB3	1:D:102:PRO:HB3	1.96	0.48
1:H:89:TRP:HB2	1:H:147:ILE:HD12	1.96	0.47
1:B:160:SER:HA	3:B:201:HOH:O	2.13	0.47
1:H:96:GLN:HE21	1:H:110:GLN:NE2	2.12	0.47
1:B:109[A]:ILE:HG12	1:B:130:TYR:OH	2.15	0.47
2:E:74:LEU:HD12	1:F:116:PHE:HE1	1.80	0.47
2:C:70:ASN:H	2:C:70:ASN:HD22	1.63	0.46
1:D:43:GLU:OE2	3:D:203:HOH:O	2.20	0.46
1:B:31:LEU:HD13	3:B:215:HOH:O	2.15	0.46
1:D:105:ARG:NH1	1:F:31:LEU:HD13	2.30	0.46
1:B:31:LEU:HD22	1:B:32:GLY:N	2.32	0.44
2:E:61:TRP:CE3	1:F:142:LYS:HA	2.52	0.44
1:B:96:GLN:HE21	1:B:110:GLN:HE21	1.66	0.44
1:H:122:ASP:OD2	3:H:202:HOH:O	2.21	0.44
1:B:114:GLY:N	1:B:125:ALA:O	2.50	0.44
1:B:17:GLU:O	3:B:205:HOH:O	2.21	0.44
1:B:76:LEU:HD22	1:B:83:TRP:CH2	2.52	0.44
2:A:70:ASN:H	2:A:70:ASN:ND2	2.16	0.44
1:F:30:LEU:HA	1:F:30:LEU:HD23	1.68	0.44
1:H:107:LYS:CD	1:H:109:ILE:HD11	2.45	0.44
1:H:46:PHE:HB3	1:H:79:TYR:HB2	2.00	0.43
1:H:142:LYS:HA	2:G:61:TRP:CE3	2.53	0.43
1:H:110:GLN:O	2:G:67:VAL:HA	2.18	0.43
1:H:65:LEU:HD22	1:H:80:CYS:SG	2.59	0.43
2:A:55:ARG:NH2	2:A:58:ASP:OD1	2.39	0.43
1:H:110:GLN:HB3	2:G:67:VAL:HG22	2.00	0.43
2:A:67:VAL:HA	1:B:110:GLN:O	2.19	0.43
1:F:42:GLN:HG2	1:F:43:GLU:HG3	2.02	0.42
2:E:59:ILE:HG23	1:F:140:LEU:HD11	2.01	0.42
1:F:161:TYR:CE2	1:F:163:SER:HB2	2.56	0.41
1:D:35:GLN:NE2	1:D:102:PRO:HA	2.35	0.41
1:F:156:ILE:HG22	1:F:157:LYS:N	2.35	0.41
1:D:143:CYS:SG	3:D:294:HOH:O	2.39	0.41
1:B:94:GLU:HG2	1:B:112:LEU:HB2	2.02	0.41
1:H:20:ASP:OD2	3:H:203:HOH:O	2.22	0.41
1:F:65:LEU:HD22	1:F:80:CYS:SG	2.60	0.41
1:D:46:PHE:HB3	1:D:79:TYR:HB2	2.03	0.41
1:F:20:ASP:OD1	3:F:203:HOH:O	2.22	0.41
1:H:107:LYS:HE3	1:H:107:LYS:HB3	1.75	0.41

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:A:70:ASN:ND2	3:A:104:HOH:O	2.53	0.40
2:A:80:GLU:HG3	3:A:136:HOH:O	2.19	0.40
2:A:61:TRP:CE3	1:B:142:LYS:HA	2.55	0.40
1:H:114:GLY:N	1:H:125:ALA:O	2.52	0.40
1:B:105:ARG:NE	3:B:220:HOH:O	2.54	0.40
1:H:105:ARG:HG2	1:H:106:ALA:N	2.36	0.40

All (7) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:A:107:HOH:O	3:A:111:HOH:O[3_444]	2.01	0.19
3:H:330:HOH:O	3:G:114:HOH:O[3_554]	2.01	0.19
3:H:307:HOH:O	3:A:139:HOH:O[1_645]	2.09	0.11
3:D:285:HOH:O	3:F:283:HOH:O[4_555]	2.13	0.07
3:G:139:HOH:O	3:G:143:HOH:O[3_554]	2.15	0.05
3:H:244:HOH:O	3:H:248:HOH:O[3_554]	2.17	0.03
3:H:320:HOH:O	3:B:309:HOH:O[3_444]	2.17	0.03

5.3 Torsion angles

5.3.1 Protein backbone

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	156/178 (88%)	151 (97%)	5 (3%)	0	100	100
1	D	145/178 (82%)	142 (98%)	3 (2%)	0	100	100
1	F	150/178 (84%)	139 (93%)	11 (7%)	0	100	100
1	H	150/178 (84%)	145 (97%)	4 (3%)	1 (1%)	26	7
2	A	36/53 (68%)	35 (97%)	1 (3%)	0	100	100
2	C	35/53 (66%)	34 (97%)	1 (3%)	0	100	100
2	E	36/53 (68%)	35 (97%)	1 (3%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	G	38/53 (72%)	38 (100%)	0	0	100	100
All	All	746/924 (81%)	719 (96%)	26 (4%)	1 (0%)	56	29

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	H	61	GLY

5.3.2 Protein sidechains ⓘ

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	119/140 (85%)	115 (97%)	4 (3%)	44	15
1	D	113/140 (81%)	111 (98%)	2 (2%)	66	40
1	F	112/140 (80%)	111 (99%)	1 (1%)	84	69
1	H	118/140 (84%)	118 (100%)	0	100	100
2	A	28/45 (62%)	27 (96%)	1 (4%)	42	13
2	C	29/45 (64%)	27 (93%)	2 (7%)	19	3
2	E	28/45 (62%)	28 (100%)	0	100	100
2	G	30/45 (67%)	30 (100%)	0	100	100
All	All	577/740 (78%)	567 (98%)	10 (2%)	66	42

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	68	THR
2	C	70	ASN
2	A	70	ASN
1	D	54	LYS
1	D	105	ARG
1	B	19	THR
1	B	31	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	54	LYS
1	B	105	ARG
1	F	34	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (12) such sidechains are listed below:

Mol	Chain	Res	Type
1	H	35	GLN
1	H	108	ASN
1	H	110	GLN
2	C	70	ASN
2	A	70	ASN
1	D	35	GLN
1	D	108	ASN
1	B	35	GLN
1	B	108	ASN
1	B	110	GLN
1	F	35	GLN
1	F	108	ASN

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

There are no ligands in this entry.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

6 Fit of model and data ⓘ

6.1 Protein, DNA and RNA chains ⓘ

In the following table, the column labelled ‘#RSRZ> 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q< 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å ²)	Q<0.9
1	B	154/178 (86%)	-0.02	4 (2%) 59 59	9, 15, 39, 58	0
1	D	148/178 (83%)	-0.15	1 (0%) 89 89	13, 24, 45, 60	0
1	F	152/178 (85%)	0.19	7 (4%) 36 34	13, 28, 56, 78	0
1	H	153/178 (85%)	-0.03	3 (1%) 68 69	8, 15, 38, 67	0
2	A	38/53 (71%)	-0.10	0 100 100	12, 20, 32, 45	0
2	C	37/53 (69%)	0.03	0 100 100	19, 29, 48, 52	0
2	E	38/53 (71%)	0.34	2 (5%) 30 28	21, 34, 62, 66	0
2	G	40/53 (75%)	-0.11	0 100 100	13, 22, 43, 45	0
All	All	760/924 (82%)	0.01	17 (2%) 65 66	8, 22, 48, 78	0

All (17) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	30	LEU	8.6
1	F	31	LEU	4.3
2	E	87	VAL	3.7
1	F	63	GLY	3.6
1	F	29	ARG	3.5
1	H	61	GLY	3.5
1	B	159	GLY	3.1
1	F	62	GLU	3.1
1	B	31	LEU	3.0
1	B	158	ASN	2.8
1	H	19	THR	2.7
1	B	30	LEU	2.6
1	H	62	GLU	2.2
1	D	33	SER	2.2
1	F	159	GLY	2.2
2	E	65	ALA	2.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	F	155	VAL	2.0

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

6.4 Ligands [i](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.

Full wwPDB X-ray Structure Validation Report ⓘ

Nov 1, 2016 – 03:28 PM JST

PDB ID : 5H4I
Title : Crystal Structure of Unlinked NS2B-NS3 Protease from Zika Virus
Deposited on : 2016-10-31
Resolution : 2.00 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report.

This report is produced by the wwPDB biocuration pipeline after annotation of the structure.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.7.1 (RC1), CSD as537be (2016)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20028320
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20028320

1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.00 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	6249 (2.00-2.00)
Clashscore	102246	7340 (2.00-2.00)
Ramachandran outliers	100387	7248 (2.00-2.00)
Sidechain outliers	100360	7247 (2.00-2.00)
RSRZ outliers	91569	6262 (2.00-2.00)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	53	<div> <div>66%</div> <div>28%</div> </div>
2	B	178	<div> <div>76%</div> <div>10%</div> <div>13%</div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	7HQ	B	201	-	-	-	X

CONFIDENTIAL VALIDATION REPORT

2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 1511 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called NS2B cofactor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	38	Total	C	N	O	S	0	0	0
			292	181	47	63	1			

- Molecule 2 is a protein called NS3 protease.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	155	Total	C	N	O	S	0	0	0
			1138	720	199	214	5			

- Molecule 3 is benzimidazol-1-ylmethanol (three-letter code: 7HQ) (formula: C₈H₈N₂O).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	B	1	Total	C	N	O	0	0
			11	8	2	1		

- Molecule 4 is ACETATE ION (three-letter code: ACT) (formula: C₂H₃O₂).

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	B	1	Total	C	O	0	0
			4	2	2		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	15	Total	O	0	0
			15	15		
5	B	51	Total	O	0	0
			51	51		

3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ($RSRZ > 2$). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: NS2B cofactor

Chain A:

- Molecule 2: NS3 protease

Chain B:

4 Data and refinement statistics

Property	Value	Source
Space group	P 43 2 2	Depositor
Cell constants a, b, c, α , β , γ	42.34 Å 42.34 Å 215.00 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	42.34 – 2.00 42.34 – 2.00	Depositor EDS
% Data completeness (in resolution range)	98.6 (42.34-2.00) 92.9 (42.34-2.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.93 (at 2.00 Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, R_{free}	0.211 , 0.244 0.211 , 0.245	Depositor DCC
R_{free} test set	686 reflections (4.93%)	DCC
Wilson B-factor (Å ²)	37.7	Xtriage
Anisotropy	0.567	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 54.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	1511	wwPDB-VP
Average B, all atoms (Å ²)	55.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.04% of the height of the origin peak. No significant pseudotranslation is detected.*

¹ Intensities estimated from amplitudes.

² Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: 7HQ, ACT

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z >5	RMSZ	# Z >5
1	A	0.24	0/296	0.47	0/401
2	B	0.26	0/1162	0.49	0/1582
All	All	0.26	0/1458	0.48	0/1983

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	292	0	269	4	0
2	B	1138	0	1111	10	0
3	B	11	0	0	0	0
4	B	4	0	3	0	0
5	A	15	0	0	1	0
5	B	51	0	0	0	0
All	All	1511	0	1383	13	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 5.

All (13) close contacts within the same asymmetric unit are listed below, sorted by their clash

magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:167:GLN:HE21	2:B:168:GLY:H	1.41	0.68
2:B:60:SER:HB2	2:B:65:LEU:HD11	1.75	0.67
1:A:70:ASN:H	1:A:70:ASN:HD22	1.43	0.64
1:A:70:ASN:H	1:A:70:ASN:ND2	2.08	0.50
2:B:35:GLN:HB3	2:B:102:PRO:HB3	1.92	0.49
1:A:73:ARG:NH1	5:A:102:HOH:O	2.46	0.47
2:B:102:PRO:HD2	2:B:131:PRO:HG2	1.96	0.46
2:B:161:TYR:CE2	2:B:163:SER:HB2	2.52	0.44
1:A:61:TRP:CH2	2:B:96:GLN:HG3	2.52	0.44
2:B:116:PHE:HB2	2:B:123:ILE:HG13	1.99	0.43
2:B:38:VAL:O	2:B:136:GLY:HA3	2.17	0.43
2:B:153:GLY:HA3	2:B:163:SER:HA	2.00	0.43
2:B:40:VAL:HG11	2:B:138:PRO:HB3	2.01	0.42

There are no symmetry-related clashes.

5.3 Torsion angles [i](#)

5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	36 / 53 (68%)	36 (100%)	0	0	100	100
2	B	153 / 178 (86%)	148 (97%)	5 (3%)	0	100	100
All	All	189 / 231 (82%)	184 (97%)	5 (3%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	31/45 (69%)	30 (97%)	1 (3%)	46	44
2	B	115/140 (82%)	110 (96%)	5 (4%)	35	30
All	All	146/185 (79%)	140 (96%)	6 (4%)	37	32

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	70	ASN
2	B	18	THR
2	B	40	VAL
2	B	123	ILE
2	B	140	LEU
2	B	158	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	70	ASN
2	B	167	GLN

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

2 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# $ Z > 2$	Counts	RMSZ	# $ Z > 2$
3	7HQ	B	201	-	10,12,12	2.63	4 (40%)	10,16,16	0.94	0
4	ACT	B	202	-	0,3,3	0.00	-	0,3,3	0.00	-

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	7HQ	B	201	-	-	0/0/2/2	0/2/2/2
4	ACT	B	202	-	-	0/0/0/0	0/0/0/0

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	201	7HQ	CAD-NAB	-4.52	1.33	1.39
3	B	201	7HQ	CAI-CAE	-4.46	1.33	1.41
3	B	201	7HQ	CAF-CAD	-3.56	1.33	1.41
3	B	201	7HQ	CAD-CAE	-3.13	1.33	1.40

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

CONFIDENTIAL VALIDATION REPORT

6 Fit of model and data [i](#)

6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled '#RSRZ > 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q < 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å ²)	Q<0.9
1	A	38/53 (71%)	0.15	0 100 100	46, 57, 81, 83	0
2	B	155/178 (87%)	0.14	2 (1%) 79 80	35, 51, 84, 106	0
All	All	193/231 (83%)	0.14	2 (1%) 84 84	35, 52, 83, 106	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	30	LEU	2.9
2	B	61	GLY	2.2

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q < 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å ²)	Q<0.9
3	7HQ	B	201	11/11	0.89	0.23	3.40	72,75,85,90	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors($\text{\AA}^2$)	Q<0.9
4	ACT	B	202	4/4	0.95	0.13	-0.12	62,63,65,67	0

6.5 Other polymers [i](#)

There are no such residues in this entry.

Full wwPDB X-ray Structure Validation Report ⓘ

May 31, 2017 – 05:40 PM JST

PDB ID : 5H6V
Title : Structure of Zika virus protease in complex with a dipeptide inhibitor
Deposited on : 2016-11-15
Resolution : 2.42 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report.

This report is produced by the wwPDB biocuration pipeline after annotation of the structure.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20029077
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20029077

1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.42 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	100719	3709 (2.44-2.40)
Clashscore	112137	4241 (2.44-2.40)
Ramachandran outliers	110173	4178 (2.44-2.40)
Sidechain outliers	110143	4179 (2.44-2.40)
RSRZ outliers	101464	3740 (2.44-2.40)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	52	 67% 6% 27%
2	B	178	 76% 10% 13%

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 1474 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Genome polyprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	38	Total	C	N	O	S	0	0	0
			292	181	47	63	1			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	44	MET	-	initiating methionine	UNP H8XX12
A	?	-	ASP	deletion	UNP H8XX12

- Molecule 2 is a protein called Genome polyprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	155	Total	C	N	O	S	0	0	0
			1138	718	197	218	5			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	0	GLY	-	expression tag	UNP H8XX12
B	29	LYS	ARG	engineered mutation	UNP H8XX12

- Molecule 3 is (S)-2-acetamido-6-amino-N-((S)-5-guanidino-1-oxopentan-2-yl)hexanamide (three-letter code: 7HS) (formula: C₁₄H₂₈N₆O₃).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	B	1	Total	C	N	O	0	0
			23	14	6	3		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	7	Total	O	0	0
			7	7		
4	B	14	Total	O	0	0
			14	14		

3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ($RSRZ > 2$). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: Genome polyprotein

Chain A:

- Molecule 2: Genome polyprotein

Chain B:

4 Data and refinement statistics

Property	Value	Source
Space group	P 43 2 2	Depositor
Cell constants a, b, c, α , β , γ	42.33 Å 42.33 Å 214.76 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	42.33 – 2.42 42.33 – 2.42	Depositor EDS
% Data completeness (in resolution range)	99.6 (42.33-2.42) 99.6 (42.33-2.42)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.15 (at 2.42 Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, R_{free}	0.191 , 0.244 0.188 , 0.245	Depositor DCC
R_{free} test set	439 reflections (5.40%)	DCC
Wilson B-factor (Å ²)	46.8	Xtriage
Anisotropy	0.591	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 53.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.45$, $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	1474	wwPDB-VP
Average B, all atoms (Å ²)	53.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.50% of the height of the origin peak. No significant pseudotranslation is detected.*

¹ Intensities estimated from amplitudes.

² Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: 7HS

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z >5	RMSZ	# Z >5
1	A	0.44	0/296	0.64	0/401
2	B	0.43	0/1162	0.62	0/1584
All	All	0.43	0/1458	0.63	0/1985

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	292	0	269	2	0
2	B	1138	0	1096	11	0
3	B	23	0	0	0	0
4	A	7	0	0	0	0
4	B	14	0	0	0	0
All	All	1474	0	1365	12	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 4.

All (12) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:24:ARG:HH11	2:B:35:GLN:HE22	1.20	0.88
2:B:101:PRO:HG2	2:B:104:GLU:HB2	1.74	0.70
2:B:64:ARG:NH1	2:B:66:ASP:OD1	2.29	0.66
2:B:90:ASP:OD2	2:B:93:SER:OG	2.13	0.66
2:B:84:LYS:O	2:B:170:ARG:NH1	2.29	0.65
1:A:55:ARG:NH1	1:A:58:ASP:OD1	2.36	0.56
2:B:86:ASP:HB3	2:B:170:ARG:HB2	1.89	0.54
2:B:170:ARG:HG2	2:B:171:GLU:H	1.79	0.48
2:B:46:PHE:HB3	2:B:79:TYR:HB2	1.98	0.46
2:B:101:PRO:HG3	2:B:107:LYS:HE2	1.99	0.44
1:A:55:ARG:NH2	2:B:20:ASP:O	2.44	0.43
2:B:97:LEU:HB3	2:B:109:ILE:HG23	2.02	0.41

There are no symmetry-related clashes.

5.3 Torsion angles [i](#)

5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	36 / 52 (69%)	35 (97%)	1 (3%)	0	100	100
2	B	153 / 178 (86%)	145 (95%)	8 (5%)	0	100	100
All	All	189 / 230 (82%)	180 (95%)	9 (5%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	31/44 (70%)	30 (97%)	1 (3%)	44 64
2	B	115/140 (82%)	112 (97%)	3 (3%)	51 71
All	All	146/184 (79%)	142 (97%)	4 (3%)	50 69

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	70	ASN
2	B	93	SER
2	B	107	LYS
2	B	118	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
2	B	35	GLN

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected

value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z > 2$	Counts	RMSZ	$\# Z > 2$
3	7HS	B	201	2	22,22,22	0.96	1 (4%)	22,26,26	1.19	1 (4%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	7HS	B	201	2	-	0/24/26/26	0/0/0/0

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	201	7HS	CAN-CAO	3.50	1.54	1.50

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	B	201	7HS	OAX-CAO-CAN	-3.24	117.58	125.15

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

6 Fit of model and data [i](#)

6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled '#RSRZ > 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q < 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ > 2		OWAB(Å ²)	Q < 0.9
1	A	38/52 (73%)	-0.09	0	100 100	41, 56, 79, 91	0
2	B	155/178 (87%)	-0.25	1 (0%)	89 87	28, 47, 83, 108	0
All	All	193/230 (83%)	-0.22	1 (0%)	90 89	28, 50, 83, 108	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	171	GLU	2.8

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q < 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å ²)	Q < 0.9
3	7HS	B	201	23/23	0.93	0.14	0.46	36,52,56,66	0

6.5 Other polymers [i](#)

There are no such residues in this entry.

CONFIDENTIAL VALIDATION REPORT

Full wwPDB X-ray Structure Validation Report ⓘ

Apr 17, 2018 – 04:46 PM JST

PDB ID : 5ZMQ
Title : Crystal structure of Zika NS3 protease with phenylacetyl-Lys-Lys-Arg-Gly-Gly-NH₂ inhibitor
Deposited on : 2018-04-05
Resolution : 1.99 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report.

This report is produced by the wwPDB biocuration pipeline after annotation of the structure.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.7.3 (157068), CSD as539be (2018)
Xtriage (Phenix) : 1.13
EDS : rb-20031021
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac : 5.8.0158
CCP4 : 7.0 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20031021

1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.99 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	111664	10189 (2.00-1.96)
Clashscore	122126	11405 (2.00-1.96)
Ramachandran outliers	120053	11281 (2.00-1.96)
Sidechain outliers	120020	11280 (2.00-1.96)
RSRZ outliers	108989	9953 (2.00-1.96)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	53	<div> <div>66%</div> <div>6%</div> <div>28%</div> </div>
1	C	53	<div> <div>66%</div> <div>.</div> <div>30%</div> </div>
1	E	53	<div> <div>66%</div> <div>6%</div> <div>28%</div> </div>
1	G	53	<div> <div>70%</div> <div>.</div> <div>26%</div> </div>
2	B	178	<div> <div>67%</div> <div>19%</div> <div>14%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	D	178
2	F	178
2	H	178
3	I	4
3	K	4

2 Entry composition i

There are 4 unique types of molecules in this entry. The entry contains 6187 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Serine protease subunit NS2B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	38	Total	C	N	O	S	0	0	0
			288	178	46	63	1			
1	C	37	Total	C	N	O	S	0	0	0
			280	175	46	58	1			
1	E	38	Total	C	N	O	S	0	0	0
			285	177	46	61	1			
1	G	39	Total	C	N	O	S	0	0	0
			293	181	47	64	1			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	44	MET	-	expression tag	UNP Q32ZE1
A	45	THR	-	expression tag	UNP Q32ZE1
C	44	MET	-	expression tag	UNP Q32ZE1
C	45	THR	-	expression tag	UNP Q32ZE1
E	44	MET	-	expression tag	UNP Q32ZE1
E	45	THR	-	expression tag	UNP Q32ZE1
G	44	MET	-	expression tag	UNP Q32ZE1
G	45	THR	-	expression tag	UNP Q32ZE1

- Molecule 2 is a protein called Serine protease NS3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	153	Total	C	N	O	S	0	0	0
			1117	706	192	215	4			
2	D	154	Total	C	N	O	S	0	0	0
			1118	709	191	214	4			
2	F	150	Total	C	N	O	S	0	0	0
			1098	698	187	209	4			
2	H	154	Total	C	N	O	S	0	0	0
			1131	718	195	214	4			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	0	GLY	-	expression tag	UNP Q32ZE1
B	107	LYS	ARG	conflict	UNP Q32ZE1
B	143	SER	CYS	conflict	UNP Q32ZE1
D	0	GLY	-	expression tag	UNP Q32ZE1
D	107	LYS	ARG	conflict	UNP Q32ZE1
D	143	SER	CYS	conflict	UNP Q32ZE1
F	0	GLY	-	expression tag	UNP Q32ZE1
F	107	LYS	ARG	conflict	UNP Q32ZE1
F	143	SER	CYS	conflict	UNP Q32ZE1
H	0	GLY	-	expression tag	UNP Q32ZE1
H	107	LYS	ARG	conflict	UNP Q32ZE1
H	143	SER	CYS	conflict	UNP Q32ZE1

- Molecule 3 is a protein (with D amino acids) called peptide PAC-DLY-DLY-DAR.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	I	4	Total	C	N	O	0	0	0
			39	26	8	5			
3	K	4	Total	C	N	O	0	0	0
			39	26	8	5			

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	37	Total	O	0	0
			37	37		
4	B	92	Total	O	0	0
			92	92		
4	C	37	Total	O	0	0
			37	37		
4	D	80	Total	O	0	0
			80	80		
4	E	33	Total	O	0	0
			33	33		
4	F	87	Total	O	0	0
			87	87		
4	G	32	Total	O	0	0
			32	32		
4	H	94	Total	O	0	0
			94	94		
4	I	6	Total	O	0	0
			6	6		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	K	1	Total	O	0	0
			1	1		

CONFIDENTIAL VALIDATION REPORT

3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ($RSRZ > 2$). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

• Molecule 1: Serine protease subunit NS2B

Chain A:

• Molecule 1: Serine protease subunit NS2B

Chain C:

• Molecule 1: Serine protease subunit NS2B

Chain E:

• Molecule 1: Serine protease subunit NS2B

Chain G:

• Molecule 2: Serine protease NS3

Chain B:

• Molecule 2: Serine protease NS3

- Molecule 2: Serine protease NS3

- Molecule 2: Serine protease NS3

- Molecule 3: peptide PAC-DLY-DLY-DAR

- Molecule 3: peptide PAC-DLY-DLY-DAR

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	60.15Å 60.33Å 214.07Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	57.91 – 1.99 57.91 – 1.99	Depositor EDS
% Data completeness (in resolution range)	98.3 (57.91-1.99) 97.7 (57.91-1.99)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.59 (at 1.98Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.171 , 0.201 0.170 , 0.202	Depositor DCC
R_{free} test set	2609 reflections (4.85%)	wwPDB-VP
Wilson B-factor (Å ²)	20.8	Xtriage
Anisotropy	0.542	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.36 , 36.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.39$, $\langle L^2 \rangle = 0.22$	Xtriage
Estimated twinning fraction	0.417 for k,h,-l	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	6187	wwPDB-VP
Average B, all atoms (Å ²)	21.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 13.96% of the height of the origin peak. No significant pseudotranslation is detected.*

¹ Intensities estimated from amplitudes.

² Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: DLY, PAC, DAR

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z >5	RMSZ	# Z >5
1	A	0.20	0/292	0.41	0/397
1	C	0.20	0/284	0.38	0/385
1	E	0.20	0/289	0.38	0/393
1	G	0.20	0/297	0.37	0/404
2	B	0.23	0/1140	0.40	0/1550
2	D	0.23	0/1141	0.40	0/1552
2	F	0.22	0/1121	0.41	0/1526
2	H	0.22	0/1153	0.40	0/1564
All	All	0.22	0/5717	0.40	0/7771

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	288	0	258	3	0
1	C	280	0	261	2	0
1	E	285	0	256	3	0
1	G	293	0	260	2	0
2	B	1117	0	1082	23	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	D	1118	0	1087	13	0
2	F	1098	0	1074	20	0
2	H	1131	0	1111	12	0
3	I	39	0	42	3	0
3	K	39	0	41	3	0
4	A	37	0	0	0	0
4	B	92	0	0	5	0
4	C	37	0	0	0	0
4	D	80	0	0	1	0
4	E	33	0	0	0	0
4	F	87	0	0	4	0
4	G	32	0	0	0	0
4	H	94	0	0	1	0
4	I	6	0	0	0	0
4	K	1	0	0	0	0
All	All	6187	0	5472	67	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 6.

All (67) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:156:ILE:O	4:B:301:HOH:O	2.11	0.69
1:E:51:MET:HG2	2:F:27:THR:HG22	1.75	0.67
2:D:64:ARG:NH1	2:D:66:ASP:OD1	2.24	0.66
2:B:101:PRO:HA	2:B:134:THR:HG22	1.78	0.63
2:F:116:PHE:HB2	2:F:123:ILE:HG23	1.79	0.63
2:B:114:GLY:N	2:B:125:ALA:O	2.32	0.62
2:B:86:ASP:HB3	2:B:170:ARG:HG3	1.86	0.57
2:H:86:ASP:OD1	2:H:86:ASP:N	2.40	0.55
2:B:93:SER:HB3	2:B:142:LYS:HE3	1.89	0.54
2:F:119:LYS:NZ	4:F:214:HOH:O	2.42	0.53
2:F:29:ARG:NH1	4:F:213:HOH:O	2.42	0.53
1:E:74:LEU:HD12	2:F:116:PHE:HE1	1.75	0.52
2:H:114:GLY:N	2:H:125:ALA:O	2.42	0.50
2:F:60:SER:HB2	2:F:65:LEU:HD11	1.92	0.50
2:B:159:GLY:N	4:B:301:HOH:O	2.45	0.50
2:H:132:ALA:HA	3:K:4:DAR:HB3	1.93	0.50
1:G:56:ALA:HB2	2:H:24:ARG:HG3	1.95	0.49
2:D:135:SER:HB2	3:I:4:DAR:C	2.43	0.49

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:F:24:ARG:NH2	2:F:104:GLU:O	2.37	0.49
2:B:161:TYR:HE2	2:B:163:SER:HB2	1.77	0.49
2:F:107:LYS:NZ	4:F:208:HOH:O	2.32	0.48
1:E:67:VAL:HG22	2:F:110:GLN:HB3	1.95	0.48
2:F:98:LEU:HB3	2:F:138:PRO:HG2	1.94	0.48
2:H:153:GLY:O	3:K:2:DLY:HB3	2.14	0.48
2:F:23:TYR:HB2	2:F:46:PHE:HE1	1.79	0.48
2:D:15:LYS:NZ	4:D:208:HOH:O	2.47	0.47
2:D:114:GLY:N	2:D:125:ALA:O	2.43	0.47
2:F:50:TRP:HE3	2:F:77:VAL:HB	1.80	0.47
1:A:62:GLU:HG2	2:B:108:ASN:HB2	1.98	0.46
1:C:56:ALA:HB3	2:D:22:VAL:HG13	1.97	0.46
2:F:39:GLY:HA3	2:F:46:PHE:CZ	2.50	0.46
2:B:20:ASP:OD1	4:B:302:HOH:O	2.21	0.46
2:B:23:TYR:HB2	2:B:46:PHE:HE1	1.81	0.46
2:B:160:SER:N	4:B:301:HOH:O	2.30	0.46
2:D:51:HIS:ND1	2:D:75:ASP:OD2	2.39	0.45
2:D:135:SER:HB2	3:I:4:DAR:HA	1.98	0.45
2:D:23:TYR:HB2	2:D:46:PHE:HE1	1.81	0.45
2:B:107:LYS:NZ	2:B:109:ILE:HD11	2.31	0.45
2:H:39:GLY:HA3	2:H:46:PHE:CZ	2.52	0.45
2:H:60:SER:HB2	2:H:65:LEU:HD11	1.99	0.45
2:B:89:TRP:HB2	2:B:147:ILE:HD12	1.99	0.45
2:F:19:THR:O	2:F:23:TYR:OH	2.29	0.44
2:H:161:TYR:CE2	2:H:163:SER:HB2	2.52	0.44
2:D:132:ALA:HA	3:I:4:DAR:HB2	2.00	0.44
2:D:22:VAL:HG21	2:D:100:VAL:HG21	1.99	0.43
2:H:132:ALA:HB2	3:K:1:PAC:H5'	2.00	0.43
2:F:159:GLY:N	4:F:209:HOH:O	2.52	0.43
2:B:39:GLY:HA3	2:B:46:PHE:CZ	2.54	0.43
2:D:39:GLY:HA3	2:D:46:PHE:CZ	2.54	0.43
2:B:116:PHE:HB2	2:B:123:ILE:HB	2.00	0.43
1:A:56:ALA:HB2	2:B:24:ARG:HG3	2.01	0.42
2:B:51:HIS:ND1	2:B:75:ASP:OD2	2.41	0.42
2:B:109:ILE:HG12	2:B:130:TYR:OH	2.19	0.42
1:G:56:ALA:HB3	2:H:22:VAL:HG13	2.00	0.42
2:F:114:GLY:N	2:F:125:ALA:O	2.40	0.41
2:F:29:ARG:HB3	2:F:29:ARG:HE	1.68	0.41
2:F:123:ILE:HD11	2:F:165:ILE:O	2.21	0.41
2:B:25:VAL:HG23	2:B:36:VAL:HG23	2.02	0.41
1:C:62:GLU:OE1	2:D:108:ASN:HB2	2.20	0.41

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:F:94:GLU:HG2	2:F:112:LEU:HB2	2.03	0.41
2:B:130:TYR:HB3	2:B:134:THR:HG21	2.02	0.41
2:H:107:LYS:NZ	4:H:202:HOH:O	2.29	0.41
2:H:58:LEU:HB2	2:H:65:LEU:HB2	2.02	0.41
1:A:60:THR:O	2:B:108:ASN:ND2	2.50	0.41
2:B:155:VAL:HG13	4:B:301:HOH:O	2.21	0.41
2:B:112:LEU:HA	2:B:113:PRO:HD3	1.94	0.40
2:D:131:PRO:HG3	2:F:104:GLU:OE2	2.21	0.40

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	36/53 (68%)	35 (97%)	1 (3%)	0	100	100
1	C	35/53 (66%)	35 (100%)	0	0	100	100
1	E	36/53 (68%)	33 (92%)	3 (8%)	0	100	100
1	G	37/53 (70%)	37 (100%)	0	0	100	100
2	B	149/178 (84%)	143 (96%)	6 (4%)	0	100	100
2	D	150/178 (84%)	145 (97%)	4 (3%)	1 (1%)	24	16
2	F	146/178 (82%)	142 (97%)	4 (3%)	0	100	100
2	H	148/178 (83%)	145 (98%)	3 (2%)	0	100	100
All	All	737/924 (80%)	715 (97%)	21 (3%)	1 (0%)	53	49

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	D	14	LYS

5.3.2 Protein sidechains ⓘ

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	30/45 (67%)	30 (100%)	0	100	100
1	C	29/45 (64%)	29 (100%)	0	100	100
1	E	29/45 (64%)	29 (100%)	0	100	100
1	G	30/45 (67%)	29 (97%)	1 (3%)	41	36
2	B	113/140 (81%)	113 (100%)	0	100	100
2	D	113/140 (81%)	113 (100%)	0	100	100
2	F	112/140 (80%)	111 (99%)	1 (1%)	81	84
2	H	115/140 (82%)	113 (98%)	2 (2%)	63	65
All	All	571/740 (77%)	567 (99%)	4 (1%)	85	88

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	F	29	ARG
1	G	50	ASP
2	H	31	LEU
2	H	135	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

6 non-standard protein/DNA/RNA residues are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and

the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
3	DLY	I	2	3	8,8,9	1.03	1 (12%)	5,8,10	1.57	1 (20%)
3	DLY	I	3	3	8,8,9	0.84	1 (12%)	5,8,10	1.54	1 (20%)
3	DAR	I	4	3	6,11,11	2.42	3 (50%)	5,13,13	0.82	0
3	DLY	K	2	3	8,8,9	0.95	1 (12%)	5,8,10	1.52	1 (20%)
3	DLY	K	3	3	8,8,9	0.73	0	5,8,10	1.26	0
3	DAR	K	4	3	6,11,11	2.39	3 (50%)	5,13,13	0.69	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	DLY	I	2	3	-	0/5/7/9	0/0/0/0
3	DLY	I	3	3	-	0/5/7/9	0/0/0/0
3	DAR	I	4	3	-	0/7/11/11	0/0/0/0
3	DLY	K	2	3	-	0/5/7/9	0/0/0/0
3	DLY	K	3	3	-	0/5/7/9	0/0/0/0
3	DAR	K	4	3	-	0/7/11/11	0/0/0/0

All (9) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	K	4	DAR	CZ-NE	-4.22	1.25	1.33
3	I	4	DAR	CZ-NE	-4.14	1.25	1.33
3	I	3	DLY	CA-C	2.15	1.53	1.50
3	K	4	DAR	CZ-NH1	2.48	1.45	1.34
3	K	2	DLY	CA-C	2.50	1.53	1.50
3	I	4	DAR	CZ-NH1	2.61	1.45	1.34
3	I	2	DLY	CA-C	2.80	1.53	1.50
3	K	4	DAR	CZ-NH2	3.20	1.44	1.32
3	I	4	DAR	CZ-NH2	3.30	1.45	1.32

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	I	2	DLY	O-C-CA	-2.58	117.88	124.96
3	K	2	DLY	O-C-CA	-2.47	118.18	124.96
3	I	3	DLY	CB-CA-C	-2.19	107.67	111.85

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	I	4	DAR	3	0
3	K	2	DLY	1	0
3	K	4	DAR	1	0

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

There are no ligands in this entry.

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

6 Fit of model and data [i](#)

6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled '#RSRZ > 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q < 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ > 2		OWAB(Å ²)	Q < 0.9
1	A	38/53 (71%)	-0.32	0	100 100	16, 21, 30, 38	0
1	C	37/53 (69%)	-0.27	0	100 100	20, 23, 28, 31	0
1	E	38/53 (71%)	-0.19	0	100 100	17, 24, 29, 33	0
1	G	39/53 (73%)	-0.16	0	100 100	18, 22, 34, 41	0
2	B	153/178 (85%)	-0.20	0	100 100	13, 19, 32, 43	0
2	D	154/178 (86%)	-0.21	4 (2%)	56 58	15, 20, 34, 45	0
2	F	150/178 (84%)	-0.22	0	100 100	14, 20, 32, 41	0
2	H	154/178 (86%)	-0.28	1 (0%)	89 90	12, 18, 29, 35	0
3	I	0/4	-	-	-	-	-
3	K	0/4	-	-	-	-	-
All	All	763/932 (81%)	-0.23	5 (0%)	87 88	12, 20, 32, 45	0

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	62	GLU	2.9
2	D	61	GLY	2.5
2	H	19	THR	2.4
2	D	13	VAL	2.3
2	D	32	GLY	2.1

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q < 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors($\text{\AA}^2$)	Q<0.9
3	DLY	K	2	9/10	0.88	0.14	17,20,28,29	0
3	DAR	K	4	12/12	0.90	0.12	19,25,31,32	0
3	DAR	I	4	12/12	0.92	0.12	12,19,31,31	0
3	DLY	I	2	9/10	0.92	0.13	22,23,25,26	0
3	DLY	I	3	9/10	0.93	0.10	17,23,25,29	0
3	DLY	K	3	9/10	0.94	0.12	11,17,23,25	0

6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

6.4 Ligands [i](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.

Preliminary Full wwPDB X-ray Structure Validation Report ⓘ

Apr 5, 2018 – 03:01 PM JST

Deposition ID : D_1300007368
PDB ID : *(not yet assigned)*

This is a Preliminary Full wwPDB X-ray Structure Validation Report.

This report is produced by the wwPDB Deposition System during initial deposition but before annotation of the structure.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	rb-20031021
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac	:	5.8.0158
CCP4	:	7.0 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20031021

1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.80 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	111664	5253 (1.80-1.80)
Clashscore	122126	6077 (1.80-1.80)
Ramachandran outliers	120053	6011 (1.80-1.80)
Sidechain outliers	120020	6010 (1.80-1.80)
RSRZ outliers	108989	5157 (1.80-1.80)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	53	
1	D	53	
1	G	53	
1	J	53	
2	B	178	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	E	178	74% 13% 13%
2	H	178	2% 69% 18% 13%
2	K	178	3% 67% 20% 13%

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 6216 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called NS2B cofactor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	38	Total	C	N	O	S	0	0	0
			291	179	46	65	1			
1	D	38	Total	C	N	O	S	0	0	0
			291	179	46	65	1			
1	G	38	Total	C	N	O	S	0	0	0
			291	179	46	65	1			
1	J	38	Total	C	N	O	S	0	0	0
			291	179	46	65	1			

- Molecule 2 is a protein called NS3 protease.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	155	Total	C	N	O	S	0	0	0
			1137	718	198	216	5			
2	E	155	Total	C	N	O	S	0	0	0
			1137	718	198	216	5			
2	H	155	Total	C	N	O	S	0	0	0
			1137	718	198	216	5			
2	K	155	Total	C	N	O	S	0	0	0
			1133	716	198	214	5			

- Molecule 3 is a ligand with the chemical component id LIG but its atom names do not match the existing wwPDB Chemical Component Dictionary definition for LIG. Consequently no firm identification of ligand chemistry can be made. Once the structure is annotated then an identification and diagram will be given here.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	C	1	Total	C	N	O	0	0
			44	28	11	5		
3	C	1	Total	C	N	O	0	0
			44	28	11	5		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	S	420	Total 420	O 420	0	0

PRELIMINARY VALIDATION REPORT

3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ($RSRZ > 2$). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: NS2B cofactor

Chain A:

- Molecule 1: NS2B cofactor

Chain D:

- Molecule 1: NS2B cofactor

Chain G:

- Molecule 1: NS2B cofactor

Chain J:

- Molecule 2: NS3 protease

Chain B:

• Molecule 2: NS3 protease

Chain E:

• Molecule 2: NS3 protease

Chain H:

• Molecule 2: NS3 protease

Chain K:

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	60.33Å 60.35Å 214.68Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	27.42 – 1.80 27.42 – 1.67	Depositor EDS
% Data completeness (in resolution range)	99.8 (27.42-1.80) 99.3 (27.42-1.67)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.41 (at 1.68Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.204 , 0.252 0.205 , 0.252	Depositor DCC
R_{free} test set	4518 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å ²)	19.5	Xtriage
Anisotropy	0.322	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 40.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.37$, $\langle L^2 \rangle = 0.19$	Xtriage
Estimated twinning fraction	0.467 for k,h,-l	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	6216	wwPDB-VP
Average B, all atoms (Å ²)	31.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 22.81 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 5.2826e-03.*

¹ Intensities estimated from amplitudes.

² Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: LIG

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z >5	RMSZ	# Z >5
1	A	0.21	0/295	0.41	0/401
1	D	0.20	0/295	0.41	0/401
1	G	0.23	0/295	0.43	0/401
1	J	0.21	0/295	0.40	0/401
2	B	0.22	0/1161	0.41	0/1581
2	E	0.22	0/1161	0.39	0/1581
2	H	0.21	0/1161	0.40	0/1581
2	K	0.21	0/1157	0.40	0/1575
All	All	0.22	0/5820	0.40	0/7922

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	291	0	260	9	0
1	D	291	0	260	1	0
1	G	291	0	260	6	0
1	J	291	0	260	7	0
2	B	1137	0	1105	28	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	E	1137	0	1105	19	0
2	H	1137	0	1105	20	0
2	K	1133	0	1095	27	0
3	C	88	0	0	7	0
4	S	420	0	0	38	0
All	All	6216	0	5450	106	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 9.

All (106) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:96:GLN:NE2	4:S:388:HOH:O	2.16	0.78
1:A:62:GLU:O	4:S:388:HOH:O	2.05	0.72
2:B:77:VAL:N	4:S:391:HOH:O	2.22	0.71
2:K:137:SER:O	4:S:53:HOH:O	2.08	0.71
2:E:49:MET:SD	4:S:119:HOH:O	2.47	0.71
2:K:24:ARG:NH2	2:K:104:GLU:O	2.25	0.70
2:E:134:THR:OG1	4:S:350:HOH:O	2.10	0.70
1:A:79:ASP:O	4:S:240:HOH:O	2.10	0.70
2:B:26:MET:SD	4:S:223:HOH:O	2.49	0.69
3:C:2:LIG:N39	4:S:344:HOH:O	2.25	0.69
2:H:24:ARG:NH1	4:S:284:HOH:O	2.28	0.66
1:J:62:GLU:HB2	2:K:108:ASN:HD22	1.60	0.66
1:A:84:PHE:O	3:C:2:LIG:N39	2.29	0.66
2:K:17:GLU:O	2:K:23:TYR:OH	2.14	0.65
2:H:164:ALA:O	4:S:295:HOH:O	2.14	0.65
2:K:43:GLU:OE1	2:K:170:ARG:NH2	2.29	0.65
2:B:35:GLN:OE1	4:S:223:HOH:O	2.15	0.65
2:K:24:ARG:NH2	2:K:101:PRO:O	2.30	0.65
2:K:158:ASN:ND2	4:S:365:HOH:O	2.31	0.64
2:K:108:ASN:ND2	4:S:59:HOH:O	2.32	0.63
2:E:150:TYR:OH	3:C:3:LIG:N34	2.32	0.62
2:K:24:ARG:NH1	2:K:35:GLN:OE1	2.31	0.62
2:B:105:ARG:NH1	4:S:235:HOH:O	2.31	0.62
3:C:2:LIG:O09	3:C:2:LIG:N33	2.32	0.62
2:K:58:LEU:HB2	2:K:65:LEU:HB2	1.82	0.62
1:A:50:ASP:N	4:S:301:HOH:O	2.33	0.61
1:J:62:GLU:O	2:K:96:GLN:NE2	2.31	0.61
2:H:114:GLY:N	2:H:125:ALA:O	2.29	0.60

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:H:64:ARG:NH1	2:H:66:ASP:OD1	2.36	0.59
2:K:139:ILE:HG13	4:S:53:HOH:O	2.02	0.59
2:H:98:LEU:HB3	2:H:138:PRO:HG2	1.84	0.59
2:B:105:ARG:NH1	4:S:243:HOH:O	2.36	0.58
2:K:98:LEU:HB3	2:K:138:PRO:HG2	1.85	0.58
1:G:73:ARG:NH1	4:S:160:HOH:O	2.35	0.58
2:B:119:LYS:NZ	4:S:128:HOH:O	2.37	0.57
2:B:114:GLY:N	2:B:125:ALA:O	2.37	0.57
2:B:19:THR:OG1	2:B:23:TYR:OH	2.23	0.56
1:J:56:ALA:HB3	2:K:22:VAL:HG13	1.87	0.56
2:B:135:SER:HB2	3:C:2:LIG:C25	2.35	0.56
2:E:131:PRO:O	2:E:134:THR:OG1	2.21	0.56
2:E:98:LEU:HB3	2:E:138:PRO:HG2	1.87	0.56
2:B:32:GLY:O	2:B:34:THR:N	2.38	0.56
2:E:114:GLY:N	2:E:125:ALA:O	2.38	0.55
2:H:58:LEU:HB2	2:H:65:LEU:HB2	1.88	0.55
1:G:70:ASN:HB2	1:G:72:PRO:HD3	1.87	0.55
1:A:50:ASP:N	4:S:275:HOH:O	2.40	0.55
2:K:143:CYS:SG	2:K:145:ARG:NH2	2.80	0.54
2:B:89:TRP:HB2	2:B:147:ILE:HD12	1.90	0.53
2:H:156:ILE:HG13	2:H:158:ASN:HB3	1.90	0.53
2:E:39:GLY:HA3	2:E:46:PHE:CZ	2.43	0.53
1:D:67:VAL:HG22	2:E:110:GLN:HB3	1.91	0.53
2:H:170:ARG:NH2	4:S:270:HOH:O	2.42	0.52
2:B:48:THR:O	4:S:391:HOH:O	2.19	0.51
2:E:51:HIS:NE2	3:C:3:LIG:O27	2.41	0.51
2:K:98:LEU:HD13	2:K:108:ASN:OD1	2.11	0.49
1:A:56:ALA:HB2	2:B:24:ARG:HG3	1.94	0.49
2:K:136:GLY:HA2	2:K:149:LEU:HD23	1.95	0.49
1:G:67:VAL:HA	2:H:110:GLN:O	2.14	0.47
2:K:96:GLN:OE1	4:S:83:HOH:O	2.20	0.47
2:B:158:ASN:HB3	2:E:129:ASP:HB3	1.96	0.47
2:B:36:VAL:HG13	4:S:337:HOH:O	2.15	0.47
1:J:79:ASP:OD1	1:J:83:ASP:N	2.42	0.47
2:H:101:PRO:HB2	2:H:104:GLU:HB2	1.98	0.46
2:B:42:GLN:O	4:S:65:HOH:O	2.20	0.46
2:K:48:THR:OG1	2:K:49:MET:N	2.49	0.46
2:K:39:GLY:HA3	2:K:46:PHE:CZ	2.51	0.46
2:E:145:ARG:NH1	4:S:342:HOH:O	2.49	0.46
2:E:24:ARG:NH2	2:E:104:GLU:O	2.28	0.46
2:B:84:LYS:HB2	4:S:65:HOH:O	2.14	0.45

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:84:LYS:O	2:B:170:ARG:NH1	2.43	0.45
2:E:105:ARG:HG2	2:E:106:ALA:H	1.81	0.45
2:K:114:GLY:N	2:K:125:ALA:O	2.35	0.45
2:B:77:VAL:HG12	4:S:391:HOH:O	2.15	0.45
2:B:59:ARG:HG2	2:B:61:GLY:H	1.82	0.45
1:G:63:LYS:O	4:S:171:HOH:O	2.21	0.44
1:J:67:VAL:HG22	2:K:110:GLN:HB3	2.00	0.44
1:A:52:TYR:CZ	2:B:26:MET:HB2	2.53	0.43
1:G:59:ILE:HD11	2:H:20:ASP:HB3	1.99	0.43
2:B:39:GLY:HA3	2:B:46:PHE:CZ	2.53	0.43
2:H:89:TRP:HB2	2:H:147:ILE:HD12	1.99	0.43
2:H:143:CYS:SG	4:S:181:HOH:O	2.62	0.43
1:G:56:ALA:HB3	2:H:22:VAL:HG13	2.00	0.43
2:K:147:ILE:O	4:S:79:HOH:O	2.22	0.42
2:H:59:ARG:NH1	4:S:352:HOH:O	2.51	0.42
1:A:70:ASN:OD1	4:S:157:HOH:O	2.21	0.42
2:H:103:GLY:N	4:S:239:HOH:O	2.45	0.42
1:J:66:GLU:HG2	1:J:68:THR:HG23	2.01	0.42
1:A:61:TRP:CD2	2:B:142:LYS:HA	2.54	0.42
2:E:98:LEU:HD11	2:E:106:ALA:HB1	2.01	0.42
2:K:71:ASP:HB3	2:K:74:GLN:HB2	2.01	0.42
2:H:18:THR:OG1	2:H:44:GLY:HA2	2.20	0.42
2:E:129:ASP:OD1	3:C:3:LIG:N33	2.53	0.42
2:E:57:ALA:N	4:S:349:HOH:O	2.51	0.42
2:B:56:ALA:O	2:B:79:TYR:OH	2.24	0.41
2:H:35:GLN:NE2	2:H:133:GLY:O	2.50	0.41
2:H:109:ILE:HG13	2:H:130:TYR:OH	2.20	0.41
2:K:97:LEU:N	2:K:109:ILE:O	2.49	0.41
2:B:24:ARG:HD3	4:S:223:HOH:O	2.20	0.41
2:B:65:LEU:HD22	2:B:80:CYS:SG	2.61	0.40
2:B:82:PRO:HG2	2:B:84:LYS:HE2	2.03	0.40
2:E:152:ASN:HB2	4:S:119:HOH:O	2.21	0.40
2:E:48:THR:OG1	2:E:49:MET:N	2.53	0.40
1:J:70:ASN:H	2:K:127:ALA:HB3	1.86	0.40
2:H:51:HIS:ND1	2:H:75:ASP:OD2	2.52	0.40
2:E:89:TRP:HB2	2:E:147:ILE:HD12	2.04	0.40
2:K:101:PRO:HA	2:K:102:PRO:HD3	1.97	0.40

There are no symmetry-related clashes.

5.3 Torsion angles [i](#)

5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	36/53 (68%)	32 (89%)	3 (8%)	1 (3%)	5	1
1	D	36/53 (68%)	34 (94%)	2 (6%)	0	100	100
1	G	36/53 (68%)	31 (86%)	5 (14%)	0	100	100
1	J	36/53 (68%)	34 (94%)	2 (6%)	0	100	100
2	B	153/178 (86%)	149 (97%)	3 (2%)	1 (1%)	24	10
2	E	153/178 (86%)	142 (93%)	10 (6%)	1 (1%)	24	10
2	H	153/178 (86%)	140 (92%)	13 (8%)	0	100	100
2	K	153/178 (86%)	143 (94%)	10 (6%)	0	100	100
All	All	756/924 (82%)	705 (93%)	48 (6%)	3 (0%)	36	22

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	33	SER
1	A	63	LYS
2	E	29	ARG

5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	31/45 (69%)	31 (100%)	0	100	100
1	D	31/45 (69%)	31 (100%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	G	31/45 (69%)	30 (97%)	1 (3%)	42	27
1	J	31/45 (69%)	29 (94%)	2 (6%)	19	6
2	B	115/140 (82%)	113 (98%)	2 (2%)	63	54
2	E	115/140 (82%)	115 (100%)	0	100	100
2	H	115/140 (82%)	114 (99%)	1 (1%)	81	77
2	K	113/140 (81%)	113 (100%)	0	100	100
All	All	582/740 (79%)	576 (99%)	6 (1%)	78	74

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	34	THR
2	B	36	VAL
1	G	70	ASN
2	H	30	LEU
1	J	60	THR
1	J	70	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
2	K	108	ASN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 2 ligands modelled in this entry, 2 could not be matched to an existing wwPDB Chemical Component Dictionary definition at this stage - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

6 Fit of model and data [i](#)

6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled '#RSRZ > 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q < 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ > 2		OWAB(Å ²)	Q < 0.9
1	A	38/53 (71%)	-0.44	0	100 100	21, 25, 32, 40	0
1	D	38/53 (71%)	-0.37	0	100 100	20, 26, 32, 36	0
1	G	38/53 (71%)	-0.07	0	100 100	28, 36, 46, 52	0
1	J	38/53 (71%)	0.10	0	100 100	33, 46, 56, 63	0
2	B	155/178 (87%)	-0.26	0	100 100	16, 22, 38, 46	0
2	E	155/178 (87%)	-0.38	0	100 100	17, 23, 33, 38	0
2	H	155/178 (87%)	-0.14	3 (1%)	66 63	25, 34, 50, 56	0
2	K	155/178 (87%)	0.02	5 (3%)	47 42	26, 40, 48, 51	0
All	All	772/924 (83%)	-0.19	8 (1%)	82 80	16, 30, 48, 63	0

All (8) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	H	34	THR	3.2
2	K	61	GLY	2.9
2	H	62	GLU	2.7
2	K	16	GLY	2.3
2	K	30	LEU	2.2
2	K	59	ARG	2.2
2	K	154	VAL	2.1
2	H	31	LEU	2.1

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q<0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors($\text{\AA}^2$)	Q<0.9
3	LIG	C	3	44/?	0.86	0.18	28,28,28,28	0
3	LIG	C	2	44/?	0.87	0.14	25,25,25,25	0

6.5 Other polymers [i](#)

There are no such residues in this entry.

Preliminary Full wwPDB X-ray Structure Validation Report ⓘ

Apr 12, 2018 – 06:10 PM JST

Deposition ID : D_1300007439
PDB ID : *(not yet assigned)*

This is a Preliminary Full wwPDB X-ray Structure Validation Report.

This report is produced by the wwPDB Deposition System during initial deposition but before annotation of the structure.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>
with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	rb-20031021
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac	:	5.8.0158
CCP4	:	7.0 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20031021

1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.00 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	111664	7193 (2.00-2.00)
Clashscore	122126	8267 (2.00-2.00)
Ramachandran outliers	120053	8166 (2.00-2.00)
Sidechain outliers	120020	8165 (2.00-2.00)
RSRZ outliers	108989	7011 (2.00-2.00)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	39	<div> <div>69%</div> <div>31%</div> </div>
1	C	39	<div> <div>69%</div> <div>31%</div> </div>
2	B	156	<div> <div>3%</div> <div>76%</div> <div>23%</div> <div>.</div> </div>
2	D	156	<div> <div>%</div> <div>74%</div> <div>25%</div> <div>.</div> </div>
3	E	38	<div> <div>89%</div> <div>11%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain	
3	G	38
4	F	154		.
4	H	154		.

2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 6160 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	39	Total	C	N	O	S	0	0	0
			297	184	48	64	1			
1	C	39	Total	C	N	O	S	0	0	0
			300	185	48	66	1			

- Molecule 2 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	156	Total	C	N	O	S	0	0	0
			1156	729	203	219	5			
2	D	156	Total	C	N	O	S	0	0	0
			1156	729	203	219	5			

- Molecule 3 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	G	38	Total	C	N	O	S	0	0	0
			295	182	47	65	1			
3	E	38	Total	C	N	O	S	0	0	0
			295	182	47	65	1			

- Molecule 4 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	H	154	Total	C	N	O	S	0	0	0
			1151	728	203	215	5			
4	F	154	Total	C	N	O	S	0	0	0
			1151	728	203	215	5			

- Molecule 5 is a ligand with the chemical component id DRG but its atom names do not match the existing wwPDB Chemical Component Dictionary definition for DRG. Consequently no firm identification of ligand chemistry can be made. Once the structure is annotated then

an identification and diagram will be given here.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	N	1	Total	C	N	O	0	0
			58	36	18	4		
5	N	1	Total	C	N	O	0	0
			58	36	18	4		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	S	243	Total	O	0	0
			243	243		

3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ($RSRZ > 2$). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1:

- Molecule 1:

- Molecule 2:

- Molecule 2:

- Molecule 3:

Chain G: 84% 16%

• Molecule 3:

Chain E: 89% 11%

• Molecule 4:

Chain H: 3% 79% 19%

• Molecule 4:

Chain F: 3% 82% 17%

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	60.42Å 60.42Å 214.87Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	42.72 – 2.00 42.72 – 1.70	Depositor EDS
% Data completeness (in resolution range)	99.5 (42.72-2.00) 85.1 (42.72-1.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.56 (at 1.70Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.257 , 0.279 0.254 , 0.265	Depositor DCC
R_{free} test set	1381 reflections (1.58%)	wwPDB-VP
Wilson B-factor (Å ²)	20.7	Xtriage
Anisotropy	0.604	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 32.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.43$, $\langle L^2 \rangle = 0.26$	Xtriage
Estimated twinning fraction	0.487 for k,h,-l	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	6160	wwPDB-VP
Average B, all atoms (Å ²)	27.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 96.24 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 6.6393e-10. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹ Intensities estimated from amplitudes.

² Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: DRG

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z > 5	RMSZ	# Z > 5
1	A	0.19	0/301	0.39	0/408
1	C	0.19	0/304	0.38	0/412
2	B	0.22	0/1180	0.47	1/1604 (0.1%)
2	D	0.21	0/1180	0.40	0/1604
3	E	0.22	0/299	0.37	0/405
3	G	0.21	0/299	0.36	0/405
4	F	0.21	0/1175	0.43	0/1595
4	H	0.22	0/1175	0.41	0/1595
All	All	0.21	0/5913	0.42	1/8028 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	61	GLY	N-CA-C	-6.21	97.59	113.10

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	297	0	271	12	0
1	C	300	0	273	9	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	1156	0	1133	26	0
2	D	1156	0	1133	25	0
3	E	295	0	271	3	0
3	G	295	0	271	5	0
4	F	1151	0	1146	23	0
4	H	1151	0	1146	26	1
5	N	116	0	0	5	0
6	S	243	0	0	8	1
All	All	6160	0	5644	114	1

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 10.

All (114) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
4:H:156:ILE:CG2	4:H:157:LYS:H	1.58	1.17
4:H:156:ILE:HG22	4:H:157:LYS:H	0.94	1.09
4:H:156:ILE:HG22	4:H:157:LYS:N	1.61	1.04
4:F:60:SER:HB3	4:F:63:GLY:HA2	1.57	0.85
4:F:62:GLU:N	4:F:63:GLY:HA3	1.95	0.80
4:H:158:ASN:ND2	4:H:160:SER:OG	2.13	0.80
2:B:40:VAL:HG21	2:B:138:PRO:HB3	1.66	0.77
4:H:158:ASN:OD1	4:H:160:SER:OG	2.02	0.76
4:H:156:ILE:CG2	4:H:157:LYS:N	2.28	0.76
4:H:17:GLU:N	4:H:60:SER:HG	1.86	0.74
4:F:164:ALA:O	6:S:246:HOH:O	2.06	0.73
4:F:17:GLU:O	6:S:251:HOH:O	2.10	0.69
4:F:60:SER:N	4:F:63:GLY:O	2.22	0.69
4:H:114:GLY:N	4:H:125:ALA:O	2.25	0.68
2:B:158:ASN:OD1	2:B:159:GLY:N	2.27	0.67
4:F:114:GLY:N	4:F:125:ALA:O	2.27	0.67
2:B:51:HIS:HA	2:B:54:LYS:HE3	1.77	0.66
1:A:59:ILE:CD1	2:B:40:VAL:HG13	2.25	0.66
4:H:27:THR:OG1	4:H:29:ARG:HG3	1.95	0.66
2:B:42:GLN:HG2	2:B:43:GLU:HG3	1.78	0.65
4:H:158:ASN:CG	4:H:160:SER:OG	2.35	0.65
4:F:17:GLU:N	4:F:60:SER:HG	1.96	0.64
2:B:35:GLN:NE2	2:B:100:VAL:O	2.31	0.63
2:D:51:HIS:HA	2:D:54:LYS:HE3	1.81	0.62
2:D:40:VAL:HG13	2:D:47:HIS:HB2	1.81	0.61

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
5:N:2:DRG:NCD	5:N:2:DRG:OBG	2.33	0.61
2:D:86:ASP:OD1	2:D:86:ASP:N	2.34	0.61
4:H:87:ALA:O	4:H:167:GLN:NE2	2.30	0.60
4:F:136:GLY:N	4:F:150:TYR:O	2.32	0.59
1:A:59:ILE:HD11	2:B:40:VAL:HG13	1.82	0.59
4:F:59:ARG:HG2	4:F:61:GLY:H	1.68	0.59
4:H:24:ARG:NH2	4:H:104:GLU:O	2.34	0.59
2:B:94:GLU:HG2	2:B:112:LEU:HB2	1.85	0.58
5:N:1:DRG:OBG	5:N:1:DRG:NCD	2.35	0.58
2:D:42:GLN:HG2	2:D:43:GLU:HG3	1.84	0.58
1:C:68:THR:OG1	1:C:69:GLY:N	2.36	0.57
2:B:86:ASP:OD1	2:B:86:ASP:N	2.38	0.56
1:A:50:ASP:CB	6:S:241:HOH:O	2.53	0.56
4:H:59:ARG:HG2	4:H:61:GLY:H	1.71	0.56
2:D:156:ILE:HD11	2:D:158:ASN:HD21	1.72	0.54
1:A:68:THR:OG1	1:A:69:GLY:N	2.39	0.54
4:H:24:ARG:NH1	4:H:35:GLN:OE1	2.39	0.53
2:D:94:GLU:HG2	2:D:112:LEU:HB2	1.90	0.52
2:D:35:GLN:NE2	2:D:100:VAL:O	2.43	0.52
4:F:24:ARG:NH2	4:F:104:GLU:O	2.41	0.52
3:E:56:ALA:HB3	4:F:22:VAL:HG13	1.91	0.52
1:C:62:GLU:HB3	1:C:65:ALA:HB2	1.92	0.51
2:D:62:GLU:N	2:D:63:GLY:HA3	2.24	0.51
2:B:49:MET:HB2	2:B:52:VAL:HG23	1.92	0.51
3:G:77:ALA:N	3:G:85:SER:O	2.35	0.51
1:C:50:ASP:N	6:S:216:HOH:O	2.44	0.51
1:C:55:ARG:NH2	1:C:58:ASP:OD1	2.45	0.50
4:H:94:GLU:O	6:S:245:HOH:O	2.19	0.50
4:F:23:TYR:HB2	4:F:46:PHE:HE1	1.77	0.49
2:D:84:LYS:O	2:D:170:ARG:NH1	2.46	0.49
3:G:62:GLU:OE2	6:S:165:HOH:O	2.20	0.49
4:F:60:SER:H	4:F:63:GLY:C	2.10	0.49
2:D:161:TYR:CE2	2:D:163:SER:HB2	2.48	0.48
1:A:56:ALA:HB3	2:B:22:VAL:HG13	1.96	0.48
1:A:56:ALA:HB2	2:B:24:ARG:HG3	1.95	0.48
4:F:59:ARG:HA	4:F:63:GLY:O	2.13	0.48
4:H:30:LEU:HB2	6:S:36:HOH:O	2.12	0.48
1:C:56:ALA:HB3	2:D:22:VAL:HG13	1.96	0.48
1:C:52:TYR:HB3	2:D:59:ARG:HB3	1.95	0.48
4:H:42:GLN:HG3	4:F:157:LYS:NZ	2.29	0.47
3:E:55:ARG:NH2	4:F:20:ASP:O	2.40	0.47

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:79:ASP:OD1	1:A:83:ASP:N	2.38	0.47
3:G:56:ALA:HB3	4:H:22:VAL:HG13	1.97	0.46
2:B:161:TYR:CE2	2:B:163:SER:HB2	2.51	0.46
2:B:38:VAL:N	2:B:135:SER:O	2.40	0.46
2:D:116:PHE:HB2	2:D:123:ILE:HB	1.97	0.46
4:H:23:TYR:HB2	4:H:46:PHE:HE1	1.80	0.46
2:D:49:MET:HB2	2:D:52:VAL:HG23	1.98	0.45
2:D:24:ARG:NH1	2:D:35:GLN:OE1	2.44	0.45
3:G:55:ARG:NH2	4:H:20:ASP:O	2.35	0.45
1:C:73:ARG:NH1	6:S:155:HOH:O	2.40	0.45
2:B:40:VAL:CG2	2:B:138:PRO:HB3	2.43	0.45
4:H:160:SER:HG	4:H:160:SER:H	1.59	0.45
2:B:23:TYR:HB2	2:B:46:PHE:HE1	1.83	0.44
1:C:73:ARG:HB3	2:D:117:LYS:HE2	1.99	0.44
2:B:133:GLY:N	5:N:1:DRG:OBM	2.45	0.44
2:D:50:TRP:HZ3	2:D:70:GLY:HA3	1.83	0.44
1:A:62:GLU:HB3	1:A:65:ALA:HB2	1.99	0.43
4:F:98:LEU:HB3	4:F:138:PRO:HG2	2.00	0.43
3:G:51:MET:HG2	4:H:27:THR:HB	1.99	0.43
2:B:116:PHE:HB2	2:B:123:ILE:HB	2.00	0.43
2:B:89:TRP:HB2	2:B:147:ILE:HD12	1.99	0.43
2:D:86:ASP:HB3	2:D:170:ARG:HB2	2.00	0.43
4:F:116:PHE:CE2	4:F:164:ALA:HB2	2.54	0.43
4:F:62:GLU:N	4:F:63:GLY:CA	2.75	0.43
4:F:101:PRO:HB2	4:F:104:GLU:HB3	2.01	0.43
2:B:24:ARG:NH1	2:B:35:GLN:OE1	2.44	0.42
4:F:101:PRO:HA	4:F:102:PRO:HD3	1.93	0.42
4:H:116:PHE:CE2	4:H:164:ALA:HB2	2.54	0.42
2:B:115:ILE:HG23	2:B:122:ASP:HB3	2.02	0.42
4:H:65:LEU:HD22	4:H:80:CYS:SG	2.60	0.42
1:A:80:GLU:O	4:H:68:TYR:OH	2.21	0.42
2:D:116:PHE:CE2	2:D:164:ALA:HB2	2.55	0.41
1:A:50:ASP:N	2:B:29:ARG:HH11	2.18	0.41
2:D:39:GLY:HA3	2:D:46:PHE:CZ	2.55	0.41
4:H:40:VAL:HG12	4:H:138:PRO:HG3	2.02	0.41
2:D:129:ASP:OD1	5:N:2:DRG:NCE	2.53	0.41
2:D:158:ASN:ND2	2:D:160:SER:OG	2.53	0.41
1:C:79:ASP:OD1	1:C:83:ASP:N	2.46	0.41
1:A:52:TYR:HB3	2:B:59:ARG:HB3	2.02	0.41
2:D:89:TRP:HB2	2:D:147:ILE:HD12	2.02	0.41
2:D:23:TYR:HB2	2:D:46:PHE:HE1	1.85	0.41

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:74:LEU:HD12	2:B:116:PHE:HE1	1.85	0.41
2:B:62:GLU:N	2:B:63:GLY:HA3	2.35	0.41
2:B:129:ASP:OD1	5:N:1:DRG:NCE	2.54	0.41
3:E:77:ALA:N	3:E:85:SER:O	2.43	0.41
4:F:138:PRO:HB2	4:F:140:LEU:HD11	2.03	0.41
2:D:112:LEU:HA	2:D:113:PRO:HD3	1.93	0.40
4:F:60:SER:HB3	4:F:63:GLY:CA	2.39	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
4:H:30:LEU:O	6:S:249:HOH:O[1_545]	2.14	0.06

5.3 Torsion angles [i](#)

5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	37/39 (95%)	35 (95%)	2 (5%)	0	100	100
1	C	37/39 (95%)	35 (95%)	2 (5%)	0	100	100
2	B	154/156 (99%)	143 (93%)	10 (6%)	1 (1%)	27	21
2	D	154/156 (99%)	143 (93%)	11 (7%)	0	100	100
3	E	36/38 (95%)	34 (94%)	2 (6%)	0	100	100
3	G	36/38 (95%)	35 (97%)	1 (3%)	0	100	100
4	F	152/154 (99%)	142 (93%)	9 (6%)	1 (1%)	24	17
4	H	152/154 (99%)	145 (95%)	6 (4%)	1 (1%)	24	17
All	All	758/774 (98%)	712 (94%)	43 (6%)	3 (0%)	36	31

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	H	156	ILE
4	F	156	ILE
2	B	156	ILE

5.3.2 Protein sidechains ⓘ

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	31/33 (94%)	31 (100%)	0	100	100
1	C	32/33 (97%)	32 (100%)	0	100	100
2	B	118/122 (97%)	117 (99%)	1 (1%)	83	87
2	D	118/122 (97%)	113 (96%)	5 (4%)	32	29
3	E	32/32 (100%)	32 (100%)	0	100	100
3	G	32/32 (100%)	32 (100%)	0	100	100
4	F	119/121 (98%)	114 (96%)	5 (4%)	32	29
4	H	119/121 (98%)	117 (98%)	2 (2%)	63	68
All	All	601/616 (98%)	588 (98%)	13 (2%)	55	58

All (13) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	158	ASN
2	D	30	LEU
2	D	40	VAL
2	D	66	ASP
2	D	98	LEU
2	D	158	ASN
4	H	40	VAL
4	H	140	LEU
4	F	29	ARG
4	F	64	ARG
4	F	129	ASP
4	F	140	LEU
4	F	156	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
2	D	158	ASN

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 2 ligands modelled in this entry, 2 could not be matched to an existing wwPDB Chemical Component Dictionary definition at this stage - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

6 Fit of model and data ⓘ

6.1 Protein, DNA and RNA chains ⓘ

In the following table, the column labelled '#RSRZ > 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q < 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ > 2	OWAB(Å ²)	Q < 0.9
1	A	39/39 (100%)	-0.14	0 100 100	22, 29, 40, 43	0
1	C	39/39 (100%)	-0.07	0 100 100	22, 31, 45, 47	0
2	B	156/156 (100%)	-0.16	4 (2%) 56 55	18, 25, 34, 41	0
2	D	156/156 (100%)	-0.14	2 (1%) 77 76	18, 26, 38, 46	0
3	E	38/38 (100%)	-0.19	0 100 100	25, 29, 41, 45	0
3	G	38/38 (100%)	-0.05	0 100 100	22, 31, 41, 44	0
4	F	154/154 (100%)	-0.06	4 (2%) 56 55	17, 26, 36, 60	0
4	H	154/154 (100%)	-0.00	5 (3%) 47 47	19, 27, 36, 50	0
All	All	774/774 (100%)	-0.10	15 (1%) 66 65	17, 27, 40, 60	0

All (15) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	F	158	ASN	4.4
2	D	158	ASN	3.2
4	H	29	ARG	3.1
4	H	64	ARG	3.1
4	H	30	LEU	3.1
2	D	31	LEU	2.9
2	B	158	ASN	2.8
4	F	63	GLY	2.7
4	H	158	ASN	2.7
4	H	156	ILE	2.4
4	F	29	ARG	2.3
2	B	31	LEU	2.2
4	F	30	LEU	2.1
2	B	40	VAL	2.1
2	B	159	GLY	2.1

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q<0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å ²)	Q<0.9
5	DRG	N	2	58/?	0.85	0.15	17,32,43,46	0
5	DRG	N	1	58/?	0.85	0.15	17,31,51,55	0

6.5 Other polymers [i](#)

There are no such residues in this entry.

Full wwPDB X-ray Structure Validation Report ⓘ

Dec 14, 2017 – 04:32 PM JST

PDB ID : 5YVU
Title : Crystal structures of unlinked full length NS3 from Dengue virus provide insights into dynamics of protease domain
Deposited on : 2017-11-27
Resolution : 2.49 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report.

This report is produced by the wwPDB biocuration pipeline after annotation of the structure.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20030345
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20030345

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.49 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	100719	3846 (2.50-2.50)
Clashscore	112137	4554 (2.50-2.50)
Ramachandran outliers	110173	4463 (2.50-2.50)
Sidechain outliers	110143	4465 (2.50-2.50)
RSRZ outliers	101464	3876 (2.50-2.50)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	B	627	<div> <div>7%</div> <div>80%</div> <div>15%</div> <div>.</div> </div>
2	A	55	<div> <div>20%</div> <div>53%</div> <div>13%</div> <div>.</div> <div>31%</div> </div>
3	I	55	<div> <div>16%</div> <div>85%</div> <div>15%</div> </div>

2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 5350 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Genome polypeptide.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	B	601	4608	2912	812	863	21	0	0	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-8	GLY	-	expression tag	UNP F8TEL4
B	-7	GLY	-	expression tag	UNP F8TEL4
B	-6	GLY	-	expression tag	UNP F8TEL4
B	-5	GLY	-	expression tag	UNP F8TEL4
B	-4	SER	-	expression tag	UNP F8TEL4
B	-3	GLY	-	expression tag	UNP F8TEL4
B	-2	GLY	-	expression tag	UNP F8TEL4
B	-1	GLY	-	expression tag	UNP F8TEL4
B	0	GLY	-	expression tag	UNP F8TEL4
B	135	ALA	SER	engineered mutation	UNP F8TEL4

- Molecule 2 is a protein called Genome polypeptide.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
2	A	38	258	159	42	57	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	41	PHE	-	expression tag	UNP F8TEL4
A	42	LYS	-	expression tag	UNP F8TEL4
A	43	PRO	-	expression tag	UNP F8TEL4
A	44	GLY	-	expression tag	UNP F8TEL4
A	45	THR	-	expression tag	UNP F8TEL4
A	46	SER	-	expression tag	UNP F8TEL4

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	47	GLY	-	expression tag	UNP F8TEL4
A	48	SER	-	expression tag	UNP F8TEL4

- Molecule 3 is a protein called Pancreatic trypsin inhibitor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	I	55	Total	C	N	O	S	0	0	0
			430	269	80	74	7			

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	B	1	Total	C	O	0	0
			6	3	3		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	42	Total	O	0	0
			42	42		
5	A	1	Total	O	0	0
			1	1		
5	I	5	Total	O	0	0
			5	5		

3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ($RSRZ > 2$). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

• Molecule 1: Genome polyprotein

• Molecule 2: Genome polyprotein

• Molecule 3: Pancreatic trypsin inhibitor

4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	53.06 Å 85.63 Å 85.51 Å 90.00° 97.95° 90.00°	Depositor
Resolution (Å)	47.70 – 2.49 47.71 – 2.49	Depositor EDS
% Data completeness (in resolution range)	88.6 (47.70-2.49) 87.9 (47.71-2.49)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.04 (at 2.48 Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.222 , 0.266 0.222 , 0.266	Depositor DCC
R_{free} test set	2308 reflections (5.00%)	DCC
Wilson B-factor (Å ²)	51.5	Xtriage
Anisotropy	0.509	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31, 46.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	5350	wwPDB-VP
Average B, all atoms (Å ²)	72.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.74% of the height of the origin peak. No significant pseudotranslation is detected.*

¹ Intensities estimated from amplitudes.

² Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: GOL

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# $ Z > 5$	RMSZ	# $ Z > 5$
1	B	0.21	0/4711	0.40	0/6394
2	A	0.20	0/261	0.33	0/358
3	I	0.23	0/440	0.46	0/590
All	All	0.21	0/5412	0.40	0/7342

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	4608	0	4461	59	0
2	A	258	0	213	5	0
3	I	430	0	412	4	0
4	B	6	0	8	1	0
5	A	1	0	0	0	0
5	B	42	0	0	1	0
5	I	5	0	0	0	0
All	All	5350	0	5094	65	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 6.

All (65) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:508:LEU:O	1:B:513:ARG:NH1	2.21	0.73
1:B:46:PHE:HB3	1:B:79:TYR:HB2	1.71	0.72
1:B:195:PRO:HB3	1:B:463:ARG:HE	1.60	0.66
1:B:399:LEU:HD12	1:B:400:THR:HG23	1.76	0.66
1:B:290:ASP:OD1	1:B:387:ARG:NH2	2.26	0.63
1:B:396:LYS:HA	1:B:399:LEU:HG	1.81	0.62
1:B:445:GLY:HA2	1:B:447:ILE:HG23	1.82	0.62
1:B:277:ASN:HD22	1:B:303:ARG:HH21	1.45	0.61
1:B:274:ARG:NH2	5:B:804:HOH:O	2.34	0.60
2:A:83:SER:OG	2:A:84:PHE:N	2.33	0.59
1:B:267:THR:OG1	1:B:538:ARG:NH1	2.33	0.59
1:B:560:ARG:NH1	1:B:613:PHE:O	2.35	0.59
3:I:7:GLU:O	3:I:43:ASN:ND2	2.36	0.57
1:B:98:LEU:HB2	1:B:138:PRO:HG2	1.88	0.54
1:B:199:LYS:HB3	4:B:701:GOL:H31	1.88	0.54
1:B:140:ILE:HD11	2:A:59:VAL:HG13	1.90	0.53
1:B:22:VAL:HG11	1:B:138:PRO:HG3	1.90	0.53
1:B:396:LYS:O	1:B:400:THR:OG1	2.27	0.52
1:B:127:THR:HG22	1:B:162:VAL:HG22	1.91	0.51
1:B:22:VAL:HG12	1:B:40:ILE:HG23	1.92	0.51
1:B:277:ASN:HD21	1:B:308:GLU:HB3	1.73	0.51
1:B:139:ILE:HD12	1:B:165:ILE:HG23	1.93	0.51
1:B:19:SER:HB3	1:B:23:TYR:OH	2.10	0.51
1:B:45:VAL:HG11	1:B:82:GLY:O	2.11	0.50
1:B:130:PHE:O	3:I:15:LYS:HE3	2.13	0.49
1:B:153:GLY:HA3	1:B:163:SER:HA	1.95	0.48
1:B:492:LYS:NZ	1:B:496:ASP:OD1	2.46	0.48
1:B:507:THR:OG1	1:B:513:ARG:NH1	2.47	0.48
1:B:116:PHE:CD2	1:B:164:ALA:HB2	2.49	0.48
2:A:68:THR:HG22	2:A:69:GLY:H	1.78	0.48
1:B:24:ARG:NH2	1:B:104:LYS:O	2.46	0.47
2:A:79:ASP:OD1	2:A:80:GLU:N	2.42	0.47
1:B:433:ILE:HG12	1:B:441:VAL:HG22	1.97	0.46
1:B:94:ASP:OD1	1:B:95:VAL:N	2.49	0.46
1:B:277:ASN:ND2	1:B:303:ARG:HH21	2.12	0.46
3:I:1:ARG:HD2	3:I:55:CYS:HB3	1.97	0.45
1:B:120:THR:HA	1:B:121:GLY:HA2	1.77	0.44
1:B:598:ALA:HA	1:B:601:TYR:CE2	2.52	0.44
1:B:89:TRP:CZ3	1:B:91:LYS:HA	2.52	0.44
1:B:597:ASP:O	1:B:600:VAL:HG22	2.16	0.44

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:324:PRO:HG2	1:B:325:PHE:CD2	2.53	0.44
1:B:409:ASP:N	1:B:409:ASP:OD1	2.50	0.44
1:B:294:VAL:HG12	1:B:508:LEU:HD22	2.00	0.43
1:B:39:GLY:HA3	1:B:46:PHE:CZ	2.53	0.43
1:B:248:LYS:HB3	1:B:272:SER:HB2	2.01	0.43
1:B:178:VAL:HG11	1:B:207:ILE:HG12	2.01	0.43
1:B:346:THR:OG1	1:B:347:GLY:N	2.51	0.43
1:B:221:LEU:HB3	1:B:262:HIS:CD2	2.54	0.43
1:B:495:LEU:HD12	1:B:498:ILE:HD12	2.01	0.43
1:B:98:LEU:HD22	1:B:108:HIS:CD2	2.54	0.43
1:B:60:HIS:CG	1:B:61:GLU:N	2.87	0.42
1:B:221:LEU:HB3	1:B:262:HIS:HD2	1.83	0.42
1:B:558:LYS:HA	1:B:558:LYS:HD3	1.84	0.42
1:B:574:GLU:HB2	1:B:607:LEU:HD21	2.00	0.42
1:B:388:LYS:HG2	1:B:388:LYS:H	1.58	0.41
1:B:24:ARG:HD3	1:B:35:GLN:OE1	2.21	0.41
1:B:243:GLN:HB2	1:B:260:MET:HB3	2.02	0.41
1:B:85:LEU:HG	1:B:167:GLN:HE22	1.86	0.41
1:B:243:GLN:HB3	1:B:268:ARG:NH2	2.36	0.41
1:B:299:TYR:O	1:B:303:ARG:HG2	2.22	0.41
1:B:118:THR:HA	2:A:76:VAL:O	2.21	0.40
3:I:35:TYR:CZ	3:I:40:ALA:HB2	2.56	0.40
1:B:393:GLU:O	1:B:396:LYS:HG2	2.20	0.40
1:B:434:LEU:HB2	1:B:440:ARG:HG3	2.04	0.40
1:B:581:ILE:HD11	1:B:611:LYS:HG2	2.04	0.40

There are no symmetry-related clashes.

5.3 Torsion angles [i](#)

5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	B	593 / 627 (95%)	559 (94%)	33 (6%)	1 (0%)	51 73

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	A	36/55 (66%)	34 (94%)	2 (6%)	0	100	100
3	I	53/55 (96%)	50 (94%)	3 (6%)	0	100	100
All	All	682/737 (92%)	643 (94%)	38 (6%)	1 (0%)	55	76

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	121	GLY

5.3.2 Protein sidechains ⓘ

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	473/524 (90%)	467 (99%)	6 (1%)	73	90
2	A	24/47 (51%)	21 (88%)	3 (12%)	5	10
3	I	44/46 (96%)	43 (98%)	1 (2%)	56	81
All	All	541/617 (88%)	531 (98%)	10 (2%)	64	86

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	7	VAL
1	B	140	ILE
1	B	254	ARG
1	B	288	PHE
1	B	401	ASP
1	B	528	GLU
2	A	68	THR
2	A	76	VAL
2	A	78	GLN
3	I	6	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	108	HIS
1	B	277	ASN
1	B	456	GLN

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z > 2$	Counts	RMSZ	$\# Z > 2$
4	GOL	B	701	-	5,5,5	0.34	0	5,5,5	0.25	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	GOL	B	701	-	-	0/4/4/4	0/0/0/0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	B	701	GOL	1	0

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

In the following table, the column labelled '#RSRZ > 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q < 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ > 2	OWAB(Å ²)	Q < 0.9
1	B	601/627 (95%)	0.43	47 (7%) 14 14	33, 64, 123, 153	0
2	A	38/55 (69%)	1.46	11 (28%) 1 0	87, 106, 120, 134	0
3	I	55/55 (100%)	0.85	9 (16%) 2 2	65, 98, 134, 141	0
All	All	694/737 (94%)	0.52	67 (9%) 8 8	33, 69, 125, 153	0

All (67) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	43	GLU	5.2
1	B	80	GLY	5.0
1	B	116	PHE	5.0
1	B	45	VAL	4.9
1	B	41	HIS	4.8
1	B	83	TRP	4.7
1	B	42	MET	4.6
1	B	40	ILE	4.6
2	A	76	VAL	4.5
1	B	21	GLY	4.4
1	B	60	HIS	4.4
1	B	8	PRO	4.4
1	B	89	TRP	4.3
3	I	48	ALA	4.3
1	B	44	GLY	4.3
2	A	87	ARG	4.2
1	B	84	ARG	4.1
2	A	86	ILE	4.1
1	B	79	TYR	3.9
1	B	85	LEU	3.7
1	B	115	LEU	3.6
1	B	46	PHE	3.6
1	B	124	GLY	3.5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	82	GLY	3.4
1	B	119	LEU	3.4
1	B	399	LEU	3.4
1	B	34	THR	3.4
3	I	30	CYS	3.3
3	I	25	ALA	3.3
2	A	60	GLN	3.2
1	B	121	GLY	3.1
1	B	118	THR	3.1
1	B	120	THR	3.1
1	B	29	GLY	3.0
1	B	165	ILE	3.0
1	B	47	HIS	3.0
1	B	67	PRO	3.0
1	B	123	ILE	2.9
1	B	65	LEU	2.8
2	A	82	GLY	2.7
2	A	74	ILE	2.7
1	B	23	TYR	2.7
3	I	31	GLN	2.7
1	B	156	THR	2.6
3	I	6	LEU	2.6
1	B	168	ALA	2.6
3	I	28	GLY	2.6
3	I	23	TYR	2.6
1	B	505	ILE	2.5
1	B	63	GLY	2.5
1	B	57	VAL	2.4
2	A	84	PHE	2.4
1	B	140	ILE	2.4
1	B	154	VAL	2.3
1	B	516	THR	2.3
3	I	21	TYR	2.3
2	A	66	ASP	2.3
1	B	581	ILE	2.2
1	B	159	GLY	2.2
1	B	114	GLY	2.2
1	B	147	ILE	2.1
1	B	155	VAL	2.1
2	A	59	VAL	2.1
2	A	53	LEU	2.1
3	I	50	ASP	2.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	A	51	LEU	2.1
1	B	59	CYS	2.0

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å ²)	Q<0.9
4	GOL	B	701	6/6	0.70	0.23	1.56	80,80,80,81	0

6.5 Other polymers [i](#)

There are no such residues in this entry.

Full wwPDB X-ray Structure Validation Report ⓘ

Dec 27, 2017 – 01:23 PM JST

PDB ID : 5YVJ
Title : Crystal structure of full length NS2B47-NS3 (gD4NS2BNS3) from Dengue virus 4 in open conformation
Deposited on : 2017-11-26
Resolution : 2.51 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report.

This report is produced by the wwPDB biocuration pipeline after annotation of the structure.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20030736
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20030736

1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.51 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	100719	3846 (2.50-2.50)
Clashscore	112137	4554 (2.50-2.50)
Ramachandran outliers	110173	4463 (2.50-2.50)
Sidechain outliers	110143	4465 (2.50-2.50)
RSRZ outliers	101464	3876 (2.50-2.50)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	B	618	6% 81% 14% 5%
2	C	47	2% 30% 68%

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 4842 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Genome polyprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	B	589	4598	2912	818	847	21	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	135	ALA	SER	engineered mutation	UNP F8TEL4

- Molecule 2 is a protein called Genome polyprotein.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
2	C	15	114	71	19	24	0	0	0

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: C₃H₈O₃).

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	B	1	Total	C	O	0	0
			6	3	3		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	120	Total	O	0	0
			120	120		
4	C	4	Total	O	0	0
			4	4		

3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ($RSRZ > 2$). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

• Molecule 1: Genome polypeptide

• Molecule 2: Genome polypeptide

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	52.78Å 86.81Å 75.90Å 90.00° 93.07° 90.00°	Depositor
Resolution (Å)	45.05 – 2.51 45.05 – 2.51	Depositor EDS
% Data completeness (in resolution range)	96.2 (45.05-2.51) 93.0 (45.05-2.51)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.29 (at 2.51Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.188 , 0.246 0.188 , 0.246	Depositor DCC
R_{free} test set	1136 reflections (5.02%)	DCC
Wilson B-factor (Å ²)	38.2	Xtriage
Anisotropy	0.592	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31, 46.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	4842	wwPDB-VP
Average B, all atoms (Å ²)	51.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.39% of the height of the origin peak. No significant pseudotranslation is detected.*

¹ Intensities estimated from amplitudes.

² Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: GOL

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\# Z > 5$	RMSZ	$\# Z > 5$
1	B	0.28	0/4703	0.46	0/6371
2	C	0.19	0/115	0.39	0/156
All	All	0.28	0/4818	0.46	0/6527

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	4598	0	4538	51	0
2	C	114	0	103	1	0
3	B	6	0	8	1	0
4	B	120	0	0	5	0
4	C	4	0	0	0	0
All	All	4842	0	4649	51	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 5.

All (51) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:387:ARG:HD3	1:B:410:ILE:HD13	1.56	0.86
1:B:43:GLU:O	1:B:84:ARG:NH1	2.20	0.74
1:B:66:GLU:HG2	1:B:67:PRO:HD2	1.69	0.74
1:B:59:CYS:SG	1:B:64:ARG:NH2	2.65	0.69
1:B:530:ARG:NH2	4:B:803:HOH:O	2.26	0.68
1:B:342:ARG:NH1	4:B:806:HOH:O	2.29	0.63
1:B:195:PRO:HB2	1:B:463:ARG:HG2	1.80	0.63
1:B:60:HIS:ND1	1:B:61:GLU:O	2.32	0.62
1:B:89:TRP:HB2	1:B:147:ILE:HD12	1.80	0.62
1:B:580:GLU:HG2	1:B:590:LYS:HG2	1.82	0.60
1:B:178:VAL:HG21	1:B:203:ILE:HG23	1.85	0.58
1:B:45:VAL:HG23	1:B:47:HIS:CE1	2.39	0.57
1:B:394:TYR:OH	1:B:398:LYS:NZ	2.37	0.56
1:B:46:PHE:HB3	1:B:79:TYR:HB2	1.87	0.56
1:B:480:LYS:HD2	1:B:480:LYS:H	1.71	0.55
1:B:178:VAL:HG21	1:B:203:ILE:CG2	2.38	0.53
1:B:322:ILE:HG23	1:B:451:PRO:HG2	1.90	0.53
1:B:25:ILE:HD12	1:B:52:VAL:HG11	1.91	0.52
1:B:89:TRP:CZ2	1:B:113:PRO:HG2	2.44	0.52
1:B:178:VAL:O	1:B:179:ASP:HB2	2.09	0.52
1:B:447:ILE:HG13	1:B:448:PRO:HD2	1.92	0.52
1:B:609:ASP:OD1	1:B:618:LYS:NZ	2.32	0.52
1:B:85:LEU:O	1:B:167:GLN:NE2	2.43	0.52
1:B:248:LYS:HE2	1:B:538:ARG:HH12	1.74	0.51
1:B:574:GLU:O	1:B:577:MET:HG2	2.10	0.51
1:B:72:VAL:O	4:B:801:HOH:O	2.19	0.50
1:B:199:LYS:HA	1:B:203:ILE:HB	1.95	0.49
1:B:40:ILE:HG12	1:B:138:PRO:HB3	1.94	0.49
1:B:502:GLU:N	1:B:502:GLU:OE1	2.46	0.48
1:B:388:LYS:HD3	1:B:605:MET:HB3	1.95	0.48
1:B:416:ASN:OD1	1:B:463:ARG:NH2	2.47	0.47
1:B:139:ILE:HD12	1:B:165:ILE:HG23	1.96	0.46
1:B:174:PRO:HD2	4:B:917:HOH:O	2.15	0.46
1:B:178:VAL:HG12	1:B:179:ASP:N	2.31	0.46
1:B:196:GLY:H	3:B:701:GOL:H11	1.81	0.45
1:B:51:HIS:ND1	1:B:75:ASP:OD2	2.40	0.45
1:B:357:LYS:HE3	1:B:402:TRP:CZ2	2.53	0.43
1:B:94:ASP:OD1	1:B:94:ASP:N	2.49	0.43
1:B:243:GLN:HB2	1:B:260:MET:HB3	2.01	0.43
1:B:96:GLN:HB2	2:C:61:TRP:CZ3	2.54	0.43
1:B:39:GLY:HA3	1:B:46:PHE:CZ	2.54	0.42
1:B:84:ARG:NH1	4:B:807:HOH:O	2.30	0.42

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:275:VAL:HA	1:B:276:PRO:HD3	1.89	0.42
1:B:402:TRP:CZ3	1:B:405:VAL:HG23	2.54	0.42
1:B:461:ILE:HD12	1:B:470:ASP:HB3	2.01	0.42
1:B:560:ARG:NH1	1:B:613:PHE:O	2.50	0.42
1:B:222:ALA:O	1:B:261:CYS:HA	2.20	0.42
1:B:453:SER:O	1:B:457:ARG:HG3	2.20	0.42
1:B:581:ILE:HD11	1:B:611:LYS:HG2	2.01	0.41
1:B:519:ILE:O	1:B:522:GLU:HG2	2.21	0.41
1:B:388:LYS:HE3	1:B:388:LYS:HB2	1.91	0.40

There are no symmetry-related clashes.

5.3 Torsion angles [i](#)

5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	583/618 (94%)	555 (95%)	25 (4%)	3 (0%)	32	53
2	C	13/47 (28%)	12 (92%)	1 (8%)	0	100	100
All	All	596/665 (90%)	567 (95%)	26 (4%)	3 (0%)	32	53

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	179	ASP
1	B	171	ILE
1	B	510	GLY

5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	481/523 (92%)	474 (98%)	7 (2%)	70	89
2	C	11/41 (27%)	11 (100%)	0	100	100
All	All	492/564 (87%)	485 (99%)	7 (1%)	71	90

All (7) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	62	THR
1	B	116	PHE
1	B	119	LEU
1	B	175	ASP
1	B	250	GLU
1	B	288	PHE
1	B	480	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

5.6 Ligand geometry ⓘ

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The

Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
3	GOL	B	701	-	5,5,5	0.33	0	5,5,5	0.38	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	GOL	B	701	-	-	0/4/4/4	0/0/0/0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	B	701	GOL	1	0

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

In the following table, the column labelled ‘#RSRZ> 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q< 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å ²)	Q<0.9
1	B	589/618 (95%)	0.36	36 (6%) 22 22	22, 47, 92, 121	0
2	C	15/47 (31%)	0.84	1 (6%) 19 19	61, 73, 96, 98	0
All	All	604/665 (90%)	0.37	37 (6%) 22 22	22, 48, 93, 121	0

All (37) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	178	VAL	7.9
1	B	173	GLU	5.5
1	B	62	THR	5.4
1	B	175	ASP	4.8
1	B	94	ASP	4.2
1	B	110	GLN	4.0
1	B	141	ASN	4.0
1	B	34	THR	3.9
1	B	119	LEU	3.9
1	B	124	GLY	3.8
1	B	152	ASN	3.3
1	B	176	TYR	3.2
1	B	42	MET	3.2
1	B	63	GLY	3.1
1	B	31	PHE	3.0
1	B	118	THR	3.0
1	B	148	GLY	2.9
1	B	32	GLY	2.8
1	B	172	GLY	2.7
1	B	111	THR	2.7
1	B	28	ARG	2.7
1	B	144	GLY	2.6
1	B	125	ALA	2.6
1	B	179	ASP	2.6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	C	49	ALA	2.5
1	B	170	ARG	2.5
1	B	112	LYS	2.5
1	B	30	LEU	2.4
1	B	345	ASN	2.3
1	B	177	GLU	2.3
1	B	196	GLY	2.2
1	B	89	TRP	2.2
1	B	140	ILE	2.1
1	B	116	PHE	2.1
1	B	399	LEU	2.1
1	B	145	LYS	2.1
1	B	164	ALA	2.0

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å ²)	Q<0.9
3	GOL	B	701	6/6	0.89	0.13	-1.80	60,61,63,65	0

6.5 Other polymers [i](#)

There are no such residues in this entry.

Full wwPDB X-ray Structure Validation Report ⓘ

Nov 28, 2017 – 03:04 PM JST

PDB ID : 5YVV
Title : Crystal structure of full length NS3 protein (gD4NS2BNS3) from DENV4 in closed conformation
Deposited on : 2017-11-27
Resolution : 3.10 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report.

This report is produced by the wwPDB biocuration pipeline after annotation of the structure.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Xtriage (Phenix) : 1.9-1692
EDS : rb-20030345
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20030345

1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.10 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	100719	1001 (3.12-3.08)
Clashscore	112137	1099 (3.12-3.08)
Ramachandran outliers	110173	1057 (3.12-3.08)
Sidechain outliers	110143	1057 (3.12-3.08)
RSRZ outliers	101464	1006 (3.12-3.08)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	B	599	<div> <div>2%</div> <div>83%</div> <div>15%</div> <div>.</div> </div>
2	A	56	<div> <div>14%</div> <div>59%</div> <div>13%</div> <div>29%</div> </div>

2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 4791 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Genome polypeptide.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	B	592	4519	2853	792	854	20	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	30	SER	LEU	engineered mutation	UNP F8TEL4
B	31	SER	PHE	engineered mutation	UNP F8TEL4

- Molecule 2 is a protein called Genome polypeptide.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	A	40	272	165	44	62	1	0	0	0

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	96	GLY	-	expression tag	UNP F8TEL4
A	97	GLY	-	expression tag	UNP F8TEL4
A	98	GLY	-	expression tag	UNP F8TEL4
A	99	GLY	-	expression tag	UNP F8TEL4
A	100	SER	-	expression tag	UNP F8TEL4
A	101	GLY	-	expression tag	UNP F8TEL4
A	102	GLY	-	expression tag	UNP F8TEL4
A	103	GLY	-	expression tag	UNP F8TEL4
A	104	GLY	-	expression tag	UNP F8TEL4

3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ($RSRZ > 2$). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

• Molecule 1: Genome polypeptide

• Molecule 2: Genome polypeptide

4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	52.92Å 88.72Å 81.30Å 90.00° 93.08° 90.00°	Depositor
Resolution (Å)	45.40 – 3.10 45.41 – 3.10	Depositor EDS
% Data completeness (in resolution range)	99.4 (45.40-3.10) 99.4 (45.41-3.10)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.21 (at 3.12Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.225 , 0.278 0.231 , 0.278	Depositor DCC
R_{free} test set	581 reflections (4.25%)	DCC
Wilson B-factor (Å ²)	73.2	Xtriage
Anisotropy	0.312	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 53.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	4791	wwPDB-VP
Average B, all atoms (Å ²)	82.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.09% of the height of the origin peak. No significant pseudotranslation is detected.*

¹ Intensities estimated from amplitudes.

² Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z >5	RMSZ	# Z >5
1	B	0.22	0/4618	0.40	0/6273
2	A	0.20	0/274	0.36	0/372
All	All	0.21	0/4892	0.40	0/6645

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	4519	0	4378	55	0
2	A	272	0	219	6	0
All	All	4791	0	4597	58	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 6.

All (58) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:26:MET:HA	1:B:35:GLN:HA	1.70	0.73
1:B:40:ILE:HD11	1:B:138:PRO:HB3	1.72	0.70

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:43:GLU:O	1:B:84:ARG:NH2	2.27	0.67
1:B:445:GLY:HA2	1:B:447:ILE:HG13	1.76	0.67
1:B:84:ARG:HB3	1:B:85:LEU:HA	1.83	0.61
1:B:495:LEU:HD22	1:B:506:PRO:HB2	1.82	0.61
1:B:127:THR:HG22	1:B:162:VAL:HG23	1.81	0.60
1:B:141:ASN:HB3	1:B:147:ILE:HD13	1.83	0.60
1:B:179:ASP:OD1	1:B:180:GLU:N	2.35	0.59
1:B:114:GLY:HA3	2:A:72:PRO:HD2	1.86	0.58
1:B:322:ILE:HG23	1:B:451:PRO:HG2	1.87	0.56
1:B:505:ILE:HD12	1:B:506:PRO:HD2	1.86	0.55
1:B:366:LYS:NZ	1:B:602:ALA:O	2.39	0.55
1:B:392:THR:O	1:B:396:LYS:NZ	2.39	0.54
2:A:70:SER:HB3	2:A:72:PRO:HD3	1.89	0.54
1:B:243:GLN:HE22	1:B:250:GLU:HG2	1.72	0.53
1:B:25:ILE:HD11	1:B:46:PHE:HE2	1.75	0.51
1:B:25:ILE:HD11	1:B:48:THR:HG21	1.93	0.51
1:B:574:GLU:O	1:B:577:MET:HG2	2.11	0.51
1:B:153:GLY:HA3	1:B:163:SER:HA	1.92	0.50
1:B:23:TYR:HA	2:A:55:LYS:HA	1.93	0.50
1:B:29:GLY:O	1:B:32:GLY:N	2.37	0.50
1:B:382:VAL:HG22	1:B:404:PHE:HB2	1.94	0.49
1:B:502:GLU:HA	1:B:504:ILE:H	1.76	0.49
1:B:97:VAL:HG22	1:B:139:ILE:HG22	1.94	0.48
1:B:130:PHE:HB3	1:B:134:THR:HG21	1.94	0.48
1:B:385:LEU:HB3	1:B:407:THR:HG23	1.95	0.48
1:B:199:LYS:NZ	1:B:285:GLU:OE2	2.48	0.47
1:B:407:THR:HG21	1:B:411:SER:OG	2.15	0.47
1:B:430:LYS:HG3	1:B:485:HIS:CD2	2.49	0.47
1:B:434:LEU:HD11	1:B:442:ILE:HD13	1.97	0.46
1:B:560:ARG:NH1	1:B:613:PHE:O	2.43	0.46
1:B:243:GLN:HB2	1:B:260:MET:HB3	1.97	0.46
2:A:70:SER:HG	2:A:71:SER:N	2.13	0.46
1:B:357:LYS:HB3	1:B:402:TRP:CZ3	2.51	0.45
1:B:195:PRO:HA	1:B:196:GLY:HA2	1.49	0.45
1:B:35:GLN:HG2	1:B:102:PRO:HB3	1.99	0.44
1:B:273:THR:HA	1:B:499:TYR:CZ	2.53	0.44
1:B:85:LEU:HB2	1:B:86:GLY:HA2	1.98	0.44
1:B:433:ILE:HG12	1:B:441:VAL:HG22	2.00	0.43
1:B:495:LEU:HD12	1:B:508:LEU:HD23	2.00	0.43
1:B:339:ILE:HG21	1:B:425:PRO:HG3	2.00	0.43
1:B:50:TRP:O	1:B:54:ARG:N	2.49	0.42

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:84:ARG:HA	1:B:84:ARG:HD3	1.92	0.42
1:B:26:MET:HG2	1:B:35:GLN:HB3	1.99	0.42
1:B:500:THR:HG21	1:B:506:PRO:HD3	2.02	0.42
1:B:154:VAL:O	1:B:162:VAL:HG12	2.20	0.42
1:B:297:ARG:NH2	1:B:512:GLU:OE1	2.53	0.42
1:B:343:SER:OG	1:B:370:ASP:OD1	2.34	0.41
2:A:61:TRP:HE1	2:A:63:GLU:HB2	1.86	0.41
1:B:212:LEU:HD12	1:B:238:LEU:HD22	2.03	0.41
1:B:461:ILE:HB	1:B:470:ASP:HB3	2.03	0.41
1:B:199:LYS:HA	1:B:203:ILE:HB	2.01	0.41
1:B:111:THR:O	1:B:113:PRO:HD3	2.21	0.41
1:B:28:ARG:HB2	1:B:29:GLY:H	1.68	0.41
1:B:448:PRO:HG2	1:B:479:LEU:HD22	2.03	0.40
1:B:141:ASN:HD22	1:B:147:ILE:HD13	1.85	0.40
1:B:110:GLN:O	2:A:67:ILE:HA	2.22	0.40

There are no symmetry-related clashes.

5.3 Torsion angles [i](#)

5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	588/599 (98%)	549 (93%)	38 (6%)	1 (0%)	51	84
2	A	36/56 (64%)	30 (83%)	6 (17%)	0	100	100
All	All	624/655 (95%)	579 (93%)	44 (7%)	1 (0%)	51	84

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	151	GLY

5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	466/510 (91%)	461 (99%)	5 (1%)	78	92
2	A	25/42 (60%)	25 (100%)	0	100	100
All	All	491/552 (89%)	486 (99%)	5 (1%)	80	92

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	28	ARG
1	B	100	ILE
1	B	120	THR
1	B	250	GLU
1	B	288	PHE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

There are no ligands in this entry.

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
2	A	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	70:SER	C	71:SER	N	5.66

6 Fit of model and data [i](#)

6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled '#RSRZ > 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q < 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ > 2	OWAB(Å ²)	Q < 0.9
1	B	592/599 (98%)	0.02	13 (2%) 62 41	25, 74, 145, 201	0
2	A	40/56 (71%)	1.01	8 (20%) 1 0	84, 133, 178, 202	0
All	All	632/655 (96%)	0.08	21 (3%) 47 24	25, 79, 150, 202	0

All (21) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	A	53	LEU	5.6
1	B	161	TYR	3.8
2	A	80	GLU	3.6
1	B	140	ILE	3.6
2	A	78	GLN	3.5
1	B	85	LEU	3.1
1	B	129	ASP	3.1
2	A	60	GLN	3.1
1	B	83	TRP	3.0
2	A	83	SER	2.9
1	B	158	SER	2.8
1	B	146	VAL	2.8
1	B	162	VAL	2.7
1	B	157	LYS	2.7
1	B	155	VAL	2.5
1	B	29	GLY	2.2
2	A	84	PHE	2.2
2	A	79	ASP	2.2
1	B	125	ALA	2.1
2	A	50	ASP	2.1
1	B	52	VAL	2.0

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

6.4 Ligands [i](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.

CONFIDENTIAL VALIDATION REPORT

Full wwPDB X-ray Structure Validation Report ⓘ

Dec 12, 2017 – 04:29 PM JST

PDB ID : 5YVW
Title : Crystal structure of full length NS3 protein (eD4NS2BNS3) from DENV4 in closed conformation
Deposited on : 2017-11-27
Resolution : 3.10 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report.

This report is produced by the wwPDB biocuration pipeline after annotation of the structure.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030345
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20030345

1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.10 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	100719	1001 (3.12-3.08)
Clashscore	112137	1099 (3.12-3.08)
Ramachandran outliers	110173	1057 (3.12-3.08)
Sidechain outliers	110143	1057 (3.12-3.08)
RSRZ outliers	101464	1006 (3.12-3.08)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	B	599	<div> <div>6%</div> <div>79%</div> <div>19%</div> <div>..</div> </div>
2	A	47	<div> <div>15%</div> <div>53%</div> <div>9%</div> <div>36%</div> <div>.</div> </div>

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 4742 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Genome polyprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	B	591	4508	2851	791	846	20	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	135	ALA	SER	engineered mutation	UNP F8TEL4

- Molecule 2 is a protein called Genome polyprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	A	30	206	128	32	45	1	0	0	0

- Molecule 3 is 4-(2-HYDROXYETHYL)-1-PIPERAZINE ETHANESULFONIC ACID (three-letter code: EPE) (formula: C₈H₁₈N₂O₄S).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	B	1	Total	C	N	O	S	0	0
			15	8	2	4	1		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	13	Total	O	0	0
			13	13		

3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ($RSRZ > 2$). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

• Molecule 1: Genome polypeptide

• Molecule 2: Genome polypeptide

4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	52.77Å 88.30Å 81.11Å 90.00° 92.34° 90.00°	Depositor
Resolution (Å)	45.27 – 3.10 59.71 – 3.00	Depositor EDS
% Data completeness (in resolution range)	99.5 (45.27-3.10) 96.1 (59.71-3.00)	Depositor EDS
R_{merge}	0.16	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.80 (at 3.01Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.207 , 0.251 0.211 , 0.246	Depositor DCC
R_{free} test set	700 reflections (5.17%)	DCC
Wilson B-factor (Å ²)	48.4	Xtriage
Anisotropy	0.565	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 52.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.046 for h,-k,-l	Xtriage
F_o, F_c correlation	0.90	EDS
Total number of atoms	4742	wwPDB-VP
Average B, all atoms (Å ²)	60.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.38% of the height of the origin peak. No significant pseudotranslation is detected.*

¹ Intensities estimated from amplitudes.

² Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: EPE

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z >5	RMSZ	# Z >5
1	B	0.21	0/4610	0.42	0/6265
2	A	0.20	0/209	0.37	0/286
All	All	0.21	0/4819	0.42	0/6551

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	B	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	196	GLY	Peptide

5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	4508	0	4350	72	1
2	A	206	0	167	3	0
3	B	15	0	17	5	0
4	B	13	0	0	0	0
All	All	4742	0	4534	73	1

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 8.

All (73) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:316:ALA:HB2	3:B:701:EPE:H62	1.62	0.81
1:B:241:ARG:NH1	1:B:250:GLU:OE1	2.16	0.79
1:B:502:GLU:HB3	1:B:503:GLY:HA2	1.71	0.73
1:B:199:LYS:NZ	3:B:701:EPE:H101	2.14	0.63
1:B:141:ASN:O	1:B:143:LYS:N	2.30	0.62
1:B:290:ASP:OD1	1:B:387:ARG:NH2	2.32	0.61
1:B:560:ARG:NH1	1:B:613:PHE:O	2.33	0.61
1:B:223:PRO:HG3	1:B:285:GLU:HG3	1.83	0.60
1:B:495:LEU:HD22	1:B:506:PRO:HB2	1.84	0.60
1:B:199:LYS:HA	1:B:203:ILE:HB	1.84	0.60
1:B:445:GLY:HA2	1:B:447:ILE:HG23	1.84	0.59
1:B:392:THR:O	1:B:396:LYS:NZ	2.35	0.59
1:B:217:ARG:NH2	1:B:277:ASN:O	2.37	0.57
1:B:139:ILE:HG13	1:B:147:ILE:HG22	1.87	0.56
1:B:179:ASP:OD1	1:B:180:GLU:N	2.37	0.55
1:B:180:GLU:HG3	1:B:214:ARG:HH22	1.72	0.55
1:B:199:LYS:HZ3	3:B:701:EPE:H101	1.69	0.55
1:B:505:ILE:HD12	1:B:506:PRO:HD2	1.89	0.54
1:B:535:GLU:OE2	1:B:539:ARG:NH2	2.40	0.54
1:B:35:GLN:HG2	1:B:102:PRO:HB3	1.91	0.53
1:B:181:ASP:HB3	1:B:187:ARG:HD2	1.91	0.53
1:B:197:ALA:O	1:B:202:ARG:NH1	2.41	0.53
1:B:130:PHE:HB3	1:B:134:THR:HG21	1.91	0.52
1:B:574:GLU:O	1:B:577:MET:HG2	2.09	0.52
1:B:199:LYS:HB3	3:B:701:EPE:H102	1.91	0.52
1:B:56:SER:OG	1:B:57:VAL:N	2.43	0.51
1:B:493:MET:O	1:B:497:ASN:ND2	2.44	0.50
1:B:388:LYS:NZ	1:B:606:ALA:HA	2.27	0.50
1:B:430:LYS:HG3	1:B:485:HIS:CD2	2.48	0.48
1:B:502:GLU:HG2	1:B:504:ILE:H	1.77	0.48

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:199:LYS:HD2	1:B:314:MET:HB3	1.95	0.48
1:B:48:THR:OG1	1:B:49:MET:N	2.36	0.48
1:B:153:GLY:HA3	1:B:163:SER:HA	1.95	0.47
1:B:97:VAL:HG13	1:B:139:ILE:HG22	1.96	0.47
1:B:141:ASN:HB2	1:B:147:ILE:HB	1.96	0.47
1:B:282:VAL:HG22	1:B:312:ILE:HD12	1.97	0.47
1:B:598:ALA:HA	1:B:601:TYR:CE2	2.49	0.47
1:B:66:GLU:O	1:B:79:TYR:HB2	2.14	0.47
1:B:113:PRO:HB3	1:B:165:ILE:HD11	1.97	0.47
1:B:243:GLN:HB3	1:B:268:ARG:NH2	2.31	0.46
2:A:68:THR:HG22	2:A:69:GLY:H	1.80	0.46
1:B:426:ARG:HH11	1:B:478:PRO:HD3	1.80	0.46
1:B:434:LEU:HD11	1:B:442:ILE:HD13	1.99	0.45
1:B:407:THR:HG21	1:B:411:SER:OG	2.16	0.45
1:B:357:LYS:HB2	1:B:419:ALA:HA	1.98	0.45
1:B:110:GLN:N	2:A:66:ASP:O	2.50	0.44
1:B:169:GLU:HA	1:B:170:ARG:HA	1.51	0.44
1:B:49:MET:HA	1:B:76:MET:HA	2.00	0.44
1:B:426:ARG:NH1	1:B:478:PRO:HD3	2.32	0.44
1:B:48:THR:HG23	1:B:77:ILE:HG13	1.98	0.44
1:B:46:PHE:HB3	1:B:79:TYR:CZ	2.53	0.44
1:B:339:ILE:HG21	1:B:425:PRO:HG3	2.00	0.43
1:B:147:ILE:HA	1:B:147:ILE:HD12	1.89	0.43
1:B:178:VAL:HG12	1:B:179:ASP:O	2.19	0.43
1:B:438:PRO:O	1:B:440:ARG:NH1	2.51	0.43
1:B:567:GLU:H	1:B:567:GLU:HG3	1.62	0.43
1:B:574:GLU:HB2	1:B:607:LEU:HD21	2.00	0.42
1:B:25:ILE:HD11	1:B:46:PHE:HE1	1.85	0.42
1:B:95:VAL:HG11	1:B:139:ILE:HD12	2.02	0.42
1:B:196:GLY:H	1:B:197:ALA:HB2	1.85	0.41
1:B:461:ILE:HD12	1:B:470:ASP:HB3	2.02	0.41
1:B:217:ARG:N	1:B:279:ASN:OD1	2.53	0.41
1:B:39:GLY:HA3	1:B:46:PHE:CE1	2.55	0.41
1:B:239:PRO:HB2	1:B:256:ILE:HG22	2.01	0.41
1:B:96:GLN:HB2	1:B:110:GLN:HA	2.03	0.41
1:B:193:LEU:HD12	1:B:193:LEU:HA	1.83	0.41
1:B:195:PRO:HG2	1:B:459:GLY:HA3	2.02	0.41
1:B:285:GLU:OE2	3:B:701:EPE:H72	2.20	0.41
1:B:70:ALA:HB3	1:B:465:PRO:HB3	2.02	0.41
1:B:110:GLN:O	2:A:67:ILE:HA	2.21	0.41
1:B:199:LYS:HE2	1:B:284:ASP:OD1	2.21	0.40

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:343:SER:OG	1:B:370:ASP:OD1	2.37	0.40
1:B:248:LYS:HE2	1:B:538:ARG:HH22	1.87	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:160:ASP:OD2	1:B:594:LYS:NZ[1_656]	2.09	0.11

5.3 Torsion angles [i](#)

5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	587/599 (98%)	534 (91%)	48 (8%)	5 (1%)	20	60
2	A	28/47 (60%)	25 (89%)	2 (7%)	1 (4%)	4	23
All	All	615/646 (95%)	559 (91%)	50 (8%)	6 (1%)	18	57

All (6) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	510	GLY
1	B	84	ARG
1	B	142	LYS
1	B	501	PRO
2	A	59	VAL
1	B	174	PRO

5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar

resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	460/509 (90%)	452 (98%)	8 (2%)	66	88
2	A	19/41 (46%)	18 (95%)	1 (5%)	26	63
All	All	479/550 (87%)	470 (98%)	9 (2%)	62	87

All (9) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	60	HIS
1	B	96	GLN
1	B	129	ASP
1	B	180	GLU
1	B	181	ASP
1	B	199	LYS
1	B	288	PHE
1	B	567	GLU
2	A	68	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

5.6 Ligand geometry ⓘ

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
3	EPE	B	701	-	15,15,15	0.82	1 (6%)	18,20,20	11.88	9 (50%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	EPE	B	701	-	-	0/9/19/19	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	701	EPE	C10-S	2.82	1.81	1.77

All (9) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	B	701	EPE	O2S-S-C10	-29.27	81.65	106.79
3	B	701	EPE	O3S-S-C10	-21.29	79.87	106.06
3	B	701	EPE	O3S-S-O1S	-14.65	77.78	111.37
3	B	701	EPE	O2S-S-O1S	-9.31	81.60	113.86
3	B	701	EPE	C7-N4-C5	2.69	118.14	111.26
3	B	701	EPE	C7-N4-C3	3.55	120.37	111.26
3	B	701	EPE	C5-N4-C3	4.30	118.61	108.87
3	B	701	EPE	O3S-S-O2S	8.68	131.27	111.37
3	B	701	EPE	O1S-S-C10	28.54	131.31	106.79

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	B	701	EPE	5	0

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

CONFIDENTIAL VALIDATION REPORT

6 Fit of model and data

6.1 Protein, DNA and RNA chains

In the following table, the column labelled '#RSRZ > 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q < 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ > 2	OWAB(Å ²)	Q < 0.9
1	B	591/599 (98%)	0.14	33 (5%) 25 11	15, 46, 129, 160	0
2	A	30/47 (63%)	1.36	7 (23%) 1 0	85, 122, 142, 145	0
All	All	621/646 (96%)	0.20	40 (6%) 20 7	15, 47, 130, 160	0

All (40) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	53	THR	5.0
1	B	59	CYS	4.5
1	B	99	ALA	3.9
1	B	154	VAL	3.5
2	A	80	GLU	3.4
1	B	160	ASP	3.4
1	B	161	TYR	3.2
1	B	172	GLY	3.2
1	B	85	LEU	3.2
1	B	48	THR	3.2
1	B	63	GLY	3.1
1	B	128	LEU	3.1
1	B	95	VAL	3.1
1	B	158	SER	3.1
1	B	77	ILE	3.0
1	B	114	GLY	3.0
2	A	70	SER	2.9
2	A	75	GLU	2.9
2	A	57	ALA	2.8
2	A	59	VAL	2.8
1	B	147	ILE	2.7
1	B	141	ASN	2.7
1	B	54	ARG	2.7
1	B	25	ILE	2.6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	35	GLN	2.6
1	B	174	PRO	2.4
1	B	60	HIS	2.4
1	B	467	GLN	2.4
1	B	68	SER	2.4
1	B	159	GLY	2.4
1	B	79	TYR	2.3
1	B	56	SER	2.3
2	A	84	PHE	2.2
1	B	146	VAL	2.2
1	B	49	MET	2.2
1	B	167	GLN	2.1
1	B	129	ASP	2.1
2	A	81	ASP	2.1
1	B	82	GLY	2.1
1	B	166	THR	2.0

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors($\text{\AA}^2$)	Q<0.9
3	EPE	B	701	15/15	0.86	0.24	0.49	48,57,61,61	0

6.5 Other polymers [i](#)

There are no such residues in this entry.

Full wwPDB X-ray Structure Validation Report ⓘ

Nov 30, 2017 – 03:45 PM JST

PDB ID : 5YVY
Title : Crystal structure of unlinked full length NS3 protein (eD4NS2BNS3) from DENV4 in closed conformation
Deposited on : 2017-11-27
Resolution : 3.20 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report.

This report is produced by the wwPDB biocuration pipeline after annotation of the structure.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Xtriage (Phenix) : 1.9-1692
EDS : rb-20030345
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20030345

1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.20 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	100719	1015 (3.22-3.18)
Clashscore	112137	1009 (3.20-3.20)
Ramachandran outliers	110173	1118 (3.22-3.18)
Sidechain outliers	110143	1117 (3.22-3.18)
RSRZ outliers	101464	1020 (3.22-3.18)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	52	<div> <div>58%</div> <div>12%</div> <div>31%</div> </div>
2	B	618	<div> <div>2%</div> <div>78%</div> <div>16%</div> <div>6%</div> </div>

2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 4687 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Genome polyprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	36	Total	C	N	O	S	0	0	0
			237	144	40	52	1			

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	96	VAL	-	expression tag	UNP F8J1V3
A	97	LYS	-	expression tag	UNP F8J1V3
A	98	THR	-	expression tag	UNP F8J1V3
A	99	GLN	-	expression tag	UNP F8J1V3
A	100	ARG	-	expression tag	UNP F8J1V3

- Molecule 2 is a protein called Genome polyprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	582	Total	C	N	O	S	0	0	0
			4450	2809	787	836	18			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	30	SER	LEU	engineered mutation	UNP F8TEL4
B	31	SER	PHE	engineered mutation	UNP F8TEL4

3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ($RSRZ > 2$). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: Genome polyprotein

Chain A:

- Molecule 2: Genome polyprotein

Chain B:

4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	52.90Å 88.77Å 81.45Å 90.00° 93.85° 90.00°	Depositor
Resolution (Å)	42.97 – 3.20 59.94 – 3.20	Depositor EDS
% Data completeness (in resolution range)	84.7 (42.97-3.20) 76.4 (59.94-3.20)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.75 (at 3.19Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.205 , 0.257 0.205 , 0.257	Depositor DCC
R_{free} test set	532 reflections (5.01%)	DCC
Wilson B-factor (Å ²)	73.3	Xtriage
Anisotropy	0.261	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.27, 57.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	4687	wwPDB-VP
Average B, all atoms (Å ²)	105.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.87% of the height of the origin peak. No significant pseudotranslation is detected.*

¹ Intensities estimated from amplitudes.

² Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z >5	RMSZ	# Z >5
1	A	0.23	0/240	0.37	0/328
2	B	0.22	0/4548	0.40	0/6179
All	All	0.22	0/4788	0.39	0/6507

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	237	0	182	5	0
2	B	4450	0	4288	57	0
All	All	4687	0	4470	59	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 6.

All (59) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:46:PHE:HB3	2:B:79:TYR:HB2	1.72	0.69
2:B:495:LEU:HD22	2:B:506:PRO:HB2	1.77	0.66

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:409:ASP:OD1	2:B:457:ARG:NH2	2.27	0.63
2:B:38:VAL:HG11	2:B:100:ILE:HD12	1.82	0.62
2:B:241:ARG:NH1	2:B:250:GLU:OE2	2.32	0.62
2:B:239:PRO:HB2	2:B:256:ILE:HG22	1.83	0.61
2:B:226:VAL:HG21	2:B:390:PHE:HE2	1.66	0.61
1:A:71:SER:OG	2:B:112:LYS:O	2.16	0.60
2:B:98:LEU:HB2	2:B:138:PRO:HG2	1.85	0.58
2:B:581:ILE:HD11	2:B:611:LYS:HD2	1.85	0.58
2:B:500:THR:HG21	2:B:506:PRO:HD3	1.86	0.57
2:B:433:ILE:HG12	2:B:441:VAL:HG22	1.85	0.57
2:B:495:LEU:HD12	2:B:508:LEU:HD23	1.88	0.56
2:B:111:THR:HG21	2:B:126:VAL:HG13	1.88	0.56
2:B:244:THR:HG22	2:B:246:ALA:H	1.71	0.56
2:B:329:ASN:HB3	2:B:463:ARG:HA	1.88	0.55
1:A:76:VAL:HG11	1:A:84:PHE:HB3	1.89	0.54
2:B:97:VAL:HG22	2:B:139:ILE:HD12	1.87	0.54
2:B:303:ARG:NH2	2:B:308:GLU:OE1	2.41	0.54
2:B:434:LEU:HD11	2:B:442:ILE:HD13	1.91	0.53
2:B:52:VAL:HG23	2:B:53:THR:HG23	1.90	0.53
2:B:109:VAL:HG21	2:B:128:LEU:HD21	1.92	0.51
2:B:322:ILE:HG23	2:B:451:PRO:HG2	1.92	0.51
2:B:210:GLU:OE1	2:B:214:ARG:NH1	2.43	0.51
1:A:76:VAL:O	2:B:118:THR:HA	2.10	0.51
2:B:505:ILE:HD12	2:B:506:PRO:HD2	1.92	0.51
2:B:35:GLN:HE21	2:B:102:PRO:HA	1.74	0.51
2:B:231:MET:O	2:B:235:LEU:N	2.44	0.50
2:B:169:GLU:OE1	2:B:169:GLU:N	2.44	0.50
1:A:81:ASP:OD1	2:B:467:GLN:NE2	2.46	0.49
2:B:382:VAL:HG22	2:B:404:PHE:HB2	1.93	0.49
1:A:85:SER:OG	1:A:86:ILE:N	2.46	0.49
2:B:385:LEU:HB3	2:B:407:THR:HG23	1.96	0.48
2:B:69:TRP:CZ2	2:B:71:ASP:HB2	2.50	0.47
2:B:139:ILE:O	2:B:146:VAL:HA	2.15	0.47
2:B:191:MET:HB3	2:B:193:LEU:HG	1.96	0.46
2:B:574:GLU:O	2:B:577:MET:HG2	2.15	0.46
2:B:243:GLN:HB2	2:B:260:MET:HB3	1.98	0.45
2:B:39:GLY:HA2	2:B:48:THR:HG23	1.97	0.45
2:B:98:LEU:HD22	2:B:108:HIS:ND1	2.33	0.44
2:B:357:LYS:HB2	2:B:419:ALA:HA	1.99	0.44
2:B:407:THR:HG21	2:B:411:SER:OG	2.18	0.43
2:B:430:LYS:HG3	2:B:485:HIS:CD2	2.54	0.43

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:294:VAL:HG22	2:B:297:ARG:NH2	2.34	0.43
2:B:96:GLN:HB2	2:B:109:VAL:O	2.18	0.43
2:B:298:GLY:HA3	2:B:495:LEU:HD11	2.00	0.43
2:B:195:PRO:HA	2:B:196:GLY:HA2	1.51	0.43
2:B:48:THR:HG22	2:B:49:MET:H	1.84	0.42
2:B:199:LYS:NZ	2:B:285:GLU:OE2	2.52	0.42
2:B:70:ALA:HB2	2:B:77:ILE:HG13	2.01	0.42
2:B:426:ARG:O	2:B:448:PRO:HA	2.20	0.41
2:B:101:GLU:HA	2:B:134:THR:HG22	2.00	0.41
2:B:484:ASP:N	2:B:484:ASP:OD1	2.52	0.41
2:B:98:LEU:HD12	2:B:138:PRO:HG2	2.02	0.41
2:B:510:GLY:O	2:B:513:ARG:HG2	2.21	0.41
2:B:50:TRP:HB3	2:B:75:ASP:OD1	2.21	0.41
2:B:447:ILE:HA	2:B:448:PRO:HD2	1.91	0.40
2:B:549:LYS:HD3	2:B:549:LYS:HA	1.86	0.40
2:B:209:ARG:O	2:B:213:LYS:HG3	2.22	0.40

There are no symmetry-related clashes.

5.3 Torsion angles [i](#)

5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	34/52 (65%)	30 (88%)	4 (12%)	0	100	100
2	B	574/618 (93%)	529 (92%)	44 (8%)	1 (0%)	51	86
All	All	608/670 (91%)	559 (92%)	48 (8%)	1 (0%)	51	86

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	501	PRO

5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	20/46 (44%)	20 (100%)	0	100	100
2	B	455/524 (87%)	448 (98%)	7 (2%)	70	90
All	All	475/570 (83%)	468 (98%)	7 (2%)	70	90

All (7) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	48	THR
2	B	85	LEU
2	B	96	GLN
2	B	288	PHE
2	B	403	ASP
2	B	495	LEU
2	B	541	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
2	B	35	GLN

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

There are no ligands in this entry.

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

CONFIDENTIAL VALIDATION REPORT

6 Fit of model and data [i](#)

6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled '#RSRZ > 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q < 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ > 2		OWAB(Å ²)	Q < 0.9
1	A	36/52 (69%)	0.37	0	100 100	125, 179, 202, 205	0
2	B	582/618 (94%)	0.03	10 (1%)	70 57	44, 95, 176, 202	0
All	All	618/670 (92%)	0.05	10 (1%)	72 59	44, 99, 184, 205	0

All (10) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	154	VAL	3.9
2	B	28	ARG	3.7
2	B	155	VAL	3.3
2	B	83	TRP	3.2
2	B	217	ARG	2.5
2	B	57	VAL	2.4
2	B	177	GLU	2.4
2	B	125	ALA	2.2
2	B	148	GLY	2.2
2	B	85	LEU	2.1

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

6.4 Ligands [i](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.

CONFIDENTIAL VALIDATION REPORT

Full wwPDB X-ray Structure Validation Report ⓘ

Dec 12, 2017 – 06:44 PM JST

PDB ID : 5YW1
Title : Crystal structure of full length NS3 protein (eD4NS2BNS3) in complex with Bovine Pancreatic Trypsin Inhibitor
Deposited on : 2017-11-28
Resolution : 2.60 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report.

This report is produced by the wwPDB biocuration pipeline after annotation of the structure.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20030345
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20030345

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.60 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	100719	2542 (2.60-2.60)
Clashscore	112137	2895 (2.60-2.60)
Ramachandran outliers	110173	2848 (2.60-2.60)
Sidechain outliers	110143	2848 (2.60-2.60)
RSRZ outliers	101464	2550 (2.60-2.60)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	B	618	<div> <div>7%</div> <div>83%</div> <div>14%</div> <div>.</div> </div>
2	A	53	<div> <div>21%</div> <div>45%</div> <div>23%</div> <div>.</div> <div>30%</div> </div>
3	I	55	<div> <div>11%</div> <div>82%</div> <div>18%</div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
4	MES	B	701	-	-	-	X
5	EPE	B	702	-	-	-	X

CONFIDENTIAL

VALIDATION REPORT

2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 5346 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Genome polyprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	B	601	Total	C	N	O	S	0	0	0
			4617	2918	812	867	20			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	30	SER	LEU	engineered mutation	UNP F8TEL4
B	31	SER	PHE	engineered mutation	UNP F8TEL4

- Molecule 2 is a protein called Genome polyprotein.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	A	37	Total	C	N	O	0	0	0
			239	149	40	50			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	96	GLY	-	expression tag	UNP F8TEL4
A	97	VAL	-	expression tag	UNP F8TEL4
A	98	LYS	-	expression tag	UNP F8TEL4
A	99	THR	-	expression tag	UNP F8TEL4
A	100	GLN	-	expression tag	UNP F8TEL4
A	101	ARG	-	expression tag	UNP F8TEL4

- Molecule 3 is a protein called PTI protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	I	55	Total	C	N	O	S	0	0	0
			432	270	80	74	8			

- Molecule 4 is 2-(N-MORPHOLINO)-ETHANESULFONIC ACID (three-letter code: MES)

(formula: $C_6H_{13}NO_4S$).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
4	B	1	Total	C	N	O	S	0	0
			12	6	1	4	1		

- Molecule 5 is 4-(2-HYDROXYETHYL)-1-PIPERAZINE ETHANESULFONIC ACID (three-letter code: EPE) (formula: $C_8H_{18}N_2O_4S$).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
5	B	1	Total	C	N	O	S	0	0
			15	8	2	4	1		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	B	27	Total	O	0	0
			27	27		
6	A	3	Total	O	0	0
			3	3		
6	I	1	Total	O	0	0
			1	1		

CONFIDENTIAL VALIDATION REPORT

3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ($RSRZ > 2$). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

• Molecule 1: Genome polypeptide

• Molecule 2: Genome polypeptide

• Molecule 3: PTI protein

4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	53.02Å 87.49Å 86.45Å 90.00° 98.25° 90.00°	Depositor
Resolution (Å)	45.00 – 2.60 45.00 – 2.60	Depositor EDS
% Data completeness (in resolution range)	99.6 (45.00-2.60) 92.1 (45.00-2.60)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.03 (at 2.61Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.223 , 0.238 0.223 , 0.237	Depositor DCC
R_{free} test set	1173 reflections (4.87%)	DCC
Wilson B-factor (Å ²)	48.6	Xtriage
Anisotropy	0.544	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31, 51.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	5346	wwPDB-VP
Average B, all atoms (Å ²)	73.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.57% of the height of the origin peak. No significant pseudotranslation is detected.*

¹ Intensities estimated from amplitudes.

² Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: EPE, MES

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z >5	RMSZ	# Z >5
1	B	0.20	0/4722	0.38	0/6413
2	A	0.25	0/242	0.60	0/333
3	I	0.22	0/443	0.39	0/594
All	All	0.21	0/5407	0.40	0/7340

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	4617	0	4451	51	0
2	A	239	0	189	14	0
3	I	432	0	414	7	0
4	B	12	0	12	0	0
5	B	15	0	17	3	0
6	A	3	0	0	0	0
6	B	27	0	0	0	0
6	I	1	0	0	0	0
All	All	5346	0	5083	66	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including

hydrogen atoms). The all-atom clashscore for this structure is 6.

All (66) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:A:62:ASP:HA	2:A:63:GLU:HB2	1.65	0.76
1:B:46:PHE:HB3	1:B:79:TYR:HB2	1.70	0.72
1:B:244:THR:HG22	1:B:246:ALA:H	1.58	0.68
2:A:84:PHE:O	3:I:39:ARG:NH2	2.29	0.66
1:B:399:LEU:HD12	1:B:400:THR:HG23	1.77	0.65
1:B:445:GLY:HA2	1:B:447:ILE:HG23	1.81	0.63
1:B:225:ARG:HD3	5:B:702:EPE:H51	1.81	0.62
3:I:27:ALA:H	3:I:28:GLY:HA2	1.64	0.62
1:B:22:VAL:HG23	2:A:56:ALA:HB3	1.83	0.60
1:B:58:ILE:HG21	2:A:53:LEU:HG	1.83	0.60
1:B:19:SER:O	1:B:41:HIS:ND1	2.30	0.60
1:B:560:ARG:NH1	1:B:613:PHE:O	2.33	0.60
1:B:434:LEU:HD23	1:B:440:ARG:HH21	1.68	0.59
2:A:76:VAL:HG12	2:A:84:PHE:HB3	1.83	0.58
1:B:416:ASN:OD1	1:B:463:ARG:NH1	2.32	0.58
2:A:62:ASP:HB3	2:A:64:MET:H	1.68	0.57
2:A:83:SER:HB3	2:A:84:PHE:HA	1.88	0.55
1:B:396:LYS:O	1:B:400:THR:OG1	2.22	0.55
2:A:85:SER:HB2	3:I:39:ARG:HH12	1.73	0.53
3:I:7:GLU:O	3:I:43:ASN:ND2	2.40	0.53
1:B:24:ARG:NH2	1:B:104:LYS:O	2.41	0.53
1:B:78:SER:HB3	1:B:83:TRP:CD1	2.43	0.53
2:A:62:ASP:HA	2:A:63:GLU:CB	2.36	0.53
1:B:248:LYS:HB3	1:B:272:SER:HB2	1.91	0.53
1:B:188:LEU:HG	1:B:311:ALA:HB3	1.92	0.52
2:A:62:ASP:HB3	2:A:64:MET:N	2.26	0.50
2:A:83:SER:HB3	2:A:84:PHE:CG	2.47	0.50
3:I:27:ALA:N	3:I:28:GLY:HA2	2.28	0.49
1:B:154:VAL:HG11	2:A:76:VAL:HG11	1.96	0.48
1:B:433:ILE:HG12	1:B:441:VAL:HG22	1.96	0.48
1:B:116:PHE:CE2	1:B:164:ALA:HB2	2.49	0.48
1:B:127:THR:HG22	1:B:162:VAL:HG22	1.94	0.48
1:B:83:TRP:HZ3	1:B:167:GLN:HA	1.79	0.47
1:B:153:GLY:HA3	1:B:163:SER:HA	1.97	0.47
3:I:35:TYR:CZ	3:I:40:ALA:HB2	2.48	0.47
1:B:140:ILE:HD13	2:A:59:VAL:HG13	1.97	0.47
5:B:702:EPE:O3S	5:B:702:EPE:N1	2.48	0.47
1:B:39:GLY:HA3	1:B:46:PHE:CZ	2.50	0.47

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:19:SER:OG	1:B:20:GLU:N	2.48	0.47
1:B:261:CYS:SG	5:B:702:EPE:H71	2.55	0.47
1:B:5:TRP:CE2	1:B:67:PRO:HB2	2.50	0.47
1:B:98:LEU:HB2	1:B:138:PRO:HG2	1.97	0.46
1:B:5:TRP:CZ3	1:B:465:PRO:HD2	2.51	0.46
1:B:141:ASN:HB3	1:B:147:ILE:HD13	1.99	0.44
1:B:139:ILE:O	1:B:147:ILE:N	2.45	0.44
1:B:276:PRO:HG2	1:B:278:TYR:CZ	2.52	0.44
1:B:430:LYS:HG3	1:B:485:HIS:CD2	2.53	0.44
1:B:574:GLU:HB2	1:B:607:LEU:HD21	2.00	0.44
1:B:116:PHE:O	2:A:75:GLU:HA	2.18	0.43
1:B:294:VAL:HG12	1:B:508:LEU:HD22	2.01	0.43
1:B:26:MET:HG2	1:B:35:GLN:HA	2.00	0.43
1:B:118:THR:C	1:B:120:THR:H	2.22	0.43
3:I:1:ARG:HA	3:I:2:PRO:HD3	1.89	0.42
1:B:41:HIS:HD2	1:B:46:PHE:HD1	1.67	0.42
1:B:388:LYS:NZ	1:B:541:ASP:OD1	2.51	0.42
1:B:93:GLU:OE1	1:B:141:ASN:HB2	2.20	0.42
1:B:221:LEU:HB3	1:B:262:HIS:CD2	2.55	0.42
1:B:346:THR:OG1	1:B:347:GLY:N	2.53	0.42
1:B:241:ARG:NH1	1:B:250:GLU:OE2	2.53	0.42
1:B:485:HIS:HB3	1:B:488:TRP:CD1	2.55	0.41
1:B:89:TRP:CZ3	1:B:91:LYS:HA	2.55	0.41
1:B:223:PRO:HD2	1:B:227:VAL:HG11	2.03	0.41
1:B:263:ALA:HA	1:B:292:CYS:HB3	2.03	0.41
1:B:495:LEU:HD12	1:B:498:ILE:HD12	2.02	0.41
1:B:598:ALA:HA	1:B:601:TYR:CE2	2.56	0.41
1:B:439:GLU:HG2	1:B:545:TRP:HZ2	1.86	0.40

There are no symmetry-related clashes.

5.3 Torsion angles [i](#)

5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	593/618 (96%)	555 (94%)	36 (6%)	2 (0%)	44	70
2	A	35/53 (66%)	25 (71%)	9 (26%)	1 (3%)	5	8
3	I	53/55 (96%)	49 (92%)	4 (8%)	0	100	100
All	All	681/726 (94%)	629 (92%)	49 (7%)	3 (0%)	38	63

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	510	GLY
2	A	81	ASP
1	B	527	GLY

5.3.2 Protein sidechains ⓘ

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	475/524 (91%)	471 (99%)	4 (1%)	85	94
2	A	19/46 (41%)	17 (90%)	2 (10%)	8	15
3	I	45/47 (96%)	44 (98%)	1 (2%)	57	81
All	All	539/617 (87%)	532 (99%)	7 (1%)	73	90

All (7) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	41	HIS
1	B	156	THR
1	B	288	PHE
1	B	426	ARG
2	A	67	ILE
2	A	85	SER
3	I	6	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

2 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
4	MES	B	701	-	12,12,12	1.98	1 (8%)	14,16,16	2.40	5 (35%)
5	EPE	B	702	-	15,15,15	0.83	1 (6%)	18,20,20	1.92	6 (33%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	MES	B	701	-	-	0/6/14/14	0/1/1/1
5	EPE	B	702	-	-	0/9/19/19	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	B	701	MES	C8-S	-6.51	1.67	1.77

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	B	702	EPE	C10-S	2.76	1.81	1.77

All (11) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	B	701	MES	C6-C5-N4	-2.78	106.21	110.11
5	B	702	EPE	C7-N4-C5	2.33	117.24	111.26
5	B	702	EPE	C7-N4-C3	2.66	118.08	111.26
4	B	701	MES	C7-N4-C3	2.74	118.28	111.26
5	B	702	EPE	O1S-S-C10	2.83	109.23	106.79
5	B	702	EPE	O3S-S-C10	3.01	109.76	106.06
4	B	701	MES	O3S-S-C8	3.03	109.78	106.06
5	B	702	EPE	O2S-S-C10	3.24	109.57	106.79
5	B	702	EPE	C5-N4-C3	4.25	118.49	108.87
4	B	701	MES	C5-N4-C3	4.41	118.86	108.87
4	B	701	MES	O1S-S-C8	5.14	111.21	106.79

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	B	702	EPE	3	0

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

In the following table, the column labelled '#RSRZ > 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q < 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ > 2	OWAB(Å ²)	Q < 0.9
1	B	601/618 (97%)	0.31	45 (7%) 15 10	30, 64, 123, 153	0
2	A	37/53 (69%)	1.48	11 (29%) 1 0	78, 106, 133, 136	0
3	I	55/55 (100%)	0.46	6 (10%) 6 4	59, 96, 128, 137	0
All	All	693/726 (95%)	0.38	62 (8%) 10 6	30, 69, 126, 153	0

All (62) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	5	TRP	12.6
1	B	84	ARG	10.4
2	A	73	ILE	5.5
1	B	40	ILE	5.2
2	A	74	ILE	5.2
2	A	59	VAL	4.7
1	B	83	TRP	4.6
2	A	65	ALA	4.5
3	I	6	LEU	4.4
1	B	114	GLY	4.2
1	B	57	VAL	4.2
1	B	119	LEU	4.1
1	B	4	LEU	3.8
1	B	143	LYS	3.7
1	B	168	ALA	3.6
1	B	118	THR	3.6
1	B	149	LEU	3.5
1	B	160	ASP	3.5
1	B	116	PHE	3.4
2	A	60	GLN	3.3
1	B	8	PRO	3.3
1	B	7	VAL	3.2
1	B	121	GLY	3.0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	47	HIS	2.9
1	B	85	LEU	2.9
1	B	94	ASP	2.9
1	B	79	TYR	2.9
1	B	123	ILE	2.8
2	A	64	MET	2.8
1	B	585	GLU	2.7
2	A	81	ASP	2.7
1	B	67	PRO	2.7
1	B	157	LYS	2.6
1	B	165	ILE	2.6
1	B	399	LEU	2.5
1	B	141	ASN	2.5
1	B	113	PRO	2.5
1	B	505	ILE	2.5
1	B	196	GLY	2.4
1	B	60	HIS	2.4
1	B	58	ILE	2.4
1	B	156	THR	2.4
1	B	62	THR	2.4
2	A	68	THR	2.3
1	B	292	CYS	2.3
1	B	115	LEU	2.2
1	B	27	GLN	2.2
1	B	33	LYS	2.2
3	I	48	SER	2.2
1	B	46	PHE	2.2
2	A	76	VAL	2.2
3	I	46	LYS	2.2
1	B	154	VAL	2.1
3	I	31	GLN	2.1
1	B	41	HIS	2.1
2	A	86	ILE	2.1
3	I	30	CYS	2.1
2	A	78	GLN	2.1
3	I	25	ALA	2.0
1	B	502	GLU	2.0
1	B	48	THR	2.0
1	B	89	TRP	2.0

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å ²)	Q<0.9
5	EPE	B	702	15/15	0.79	0.30	3.84	62,75,193,193	0
4	MES	B	701	12/12	0.83	0.31	2.48	64,69,104,105	0

6.5 Other polymers [i](#)

There are no such residues in this entry.