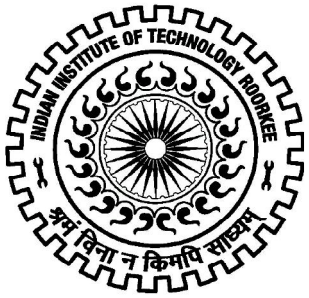
FUNCTIONAL CHARACTERIZATION OF ALPHAVIRAL PROTEINS INVOLVED IN REPLICATION

THESIS

by

MANJU NARWAL



DEPARTMENT OF BIOTECHNOLOGY
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE – 247 667 (INDIA)
JULY, 2015

FUNCTIONAL CHARACTERIZATION OF ALPHAVIRAL PROTEINS INVOLVED IN REPLICATION

A THESIS
Submitted in partial fulfilment of the requirements for the award of the degree of

DOCTOR OF PHILOSOPHY

in

BIOTECHNOLOGY

by

MANJU NARWAL

DEPARTMENT OF BIOTECHNOLOGY
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE – 247 667 (INDIA)
JULY, 2015

INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled "**FUNCTIONAL CHARACTERISATION OF ALPHAVIRAL PROTEINS INVOLVED IN REPLICATION**'' in partial fulfilment of the requirements for the award of the Degree of Doctor of Philosophy and submitted in the Department of Biotechnology of the Indian Institute of Technology Roorkee, Roorkee is an authentic record of my own work carried out during a period from December, 2009 to July, 2015 under the supervision of Dr. Shailly Tomar, Associate Professor, Department of Biotechnology, Indian Institute of Technology Roorkee, Roorkee.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other Institute.

(MANJU NARWAL)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Dated: July, 2015 (SHAILLY TOMAR)

Supervisor

ABSTRACT

Alphavirus are positive sense RNA virus which belongs to *Togaviridae* family of virus classification. These viruses have wide geographical distribution naming New world and Old world. They have a wide range of hosts ranging from humans, equine to birds and main transmission in the enzootic and epizootic cycle is assisted by *Culex* mosquito species. Other arthropods like ticks, mites, lice uphold the alphaviral replication and also act as carrier vectors. Members of this *genus* are capable of causing various range of infections from arthritis to encephalitis with high rate of mortality. However the treatment in case of alphavirus is mostly symptomatic in absence of any effective drug or vaccine. In the initial years, early 20th century when the early cases of infection are reported the alphavirus were found to be confined to certain demographics only. But with time, members of this *genus* have evolved and now have spread and covered major parts of the seven continents. And in the light of this evolution along with the absence of any therapy against this, it has become imperative to discover inhibitory drug molecules. For this purpose different crucial mechanism vital for the survival of the virus are to be targeted.

For these reasons, we have targeted alphavirus replication enzyme nsP2 (non structural protein 2) which plays important roles at the different steps of the virus replication. The N-terminal region of the protein is a helicase domain, and the C-terminal region possessing proteolytic activity is a papain-like cysteine protease. The C-terminal region is crucial for the processing of nsP1234 polyprotein, which on cleavage form individual proteins. These nsP's are part of different replication complexes which virus forms during the course of replication. Also, this C-terminal region is important for the synthesis of 26S sub-genomic RNA as it is discovered to have interactions with the promoter region for sub-genomic RNA on the genomic RNA. The protein is also crucial for inactivating the host response like IFN's. Based on these observations we have considered the C-terminal protease an ideal target for our study. By applying the biophysical and biochemical strategies we were able to decipher new features of this protein.

From three different alphaviruses viz. AURA, CHIKV and SINV, the C-terminal region was cloned in an expression vector. In case of AURAV, we have cloned and purified the protease using Ion exchange and size exclusion chromatography. The protein purification was optimized to achieve better yields of the protein for crystallization purposes. However, we were not successful in crystallizing the protein. But we have developed a simple yet reliable assay based on the application of modified β-galactosidase protein as the substrate. The

different sites at which nsP2 protease cleaves were inserted at optimum position inside the β-galactosidase gene. And then the proteolytic activity was checked on using this modified gene in presence of nsP2.

SINV nsP2 was also cloned in pET series of expression vector and was purified using affinity and size exclusion chromatography. However in case of SINV, it was difficult during the initial stages to get a soluble protein. Various modifications in the inducer concentration to temperature did not yield any success in this case. But by optimizing media conditions we were able to get pure soluble protein. The protein was purified and crystallized using sitting drop method of vapor diffusion. But we were not able to further improve the crystal quality to get diffraction quality crystals.

Protease domain from CHIKV was also cloned and purified and in this case was successfully crystallized to achieve diffraction quality crystals. The protein was crystallized with 5 mg/mL concentrations using sitting drop vapor diffusion method. The crystal was diffracted and structure was solved up to 2.6 Å. This protein is monomer in solution conditions but it was observed to be a tetramer in the crystal form. On analysis of the structure it was found to be bound to different glycerol molecules at certain positions. Also the active site region was found to be closed in comparison to the VEEV (PDB: 2HWK) and nsP23zbd SINV (PDB: 4GUA). Large amount of water molecules were found to be near the substrate binding regions of the protein. Different regions (especially active site) of high temperature factors were observed which showed that there is flexibility and chances of modifications in these regions.

Also, TNBS based activity assay was also developed in case of nsP2CHIKV. The assay was used for understanding the kinetics of nsP2 for its different substrate sites.

All these studies have shown that nsP2 protease is a unique and important protease. The protein cloned and expressed heterologously is stable and active. The protein shows preference for nsP3/4 site in comparison to other sites showing that other nsP's might be necessary for cleavage at nsP1/2 and nsP2/3 site. The information from this study might help in the understanding of the mechanism of this preference of nsP2 as well as the structure information from nsP2CHIKV would help in structure based screening of inhibitory molecules. Not only this, structure information regarding the environment of the active site, conformation and properties of active and substrate site residues would help in designing the

inhibitory molecules against nsP2 protease. This might contribute in the inhibition of the infection of the alphavirus members.

To my family & friends.

It was a great journey,

thank you for travelling with me!

Acknowledgement

I would like to express my deep gratitude to all the wonderful people I had the opportunity to work with during my PhD work. For the beginning of this story thanks belongs to Dr. Shailly Tomar, who believed in me and took me as a PhD student, a novice with no experience of the techniques to be used during this work. I also want to thank her, for her constant and untiring support for developing my scientific skills. I would like to express my gratitude for the discussions and encouragement, which have been highly appreciated during these years.

I would like to thank former and present members of the Department of biotechnology, Prof. H.S. Dhaliwal, Prof. Ritu Barthwal, Prof. R.P Singh and Prof. G.S. Randhawa, for creating such an excellent environment for scientific research, and for being available for us researchers. I have enjoyed a lot being a member of department of Biotechnology at IIT roorkee. It has offered an inspiring network of people, excellent courses and great meetings, as well as very much appreciated monetary support. I want to express my gratitude to the Head of the Biotechnology department, Prof. R. Prasad, who has always been helpful and available and for always having an open door and encouraging attitude.

I would like to thank Dr. Pravindra Kumar for a fruitful collaboration and easing the access to the XRD facility at the Institute Instrumentation Centre. Also big thanks to Dr. A.K. Sharma for the critical reading of the thesis manuscript and useful comments and suggestions, which helped to improve the manuscript to become this book we see now. For the follow up of my thesis work, thanks belongs to Dr. Shivendra and Dr. Aditya, who have created a nice and relaxed environment for scientific discussion and given me a lot of useful advice.

Last but not the least I want to thank my family and my friends. The biggest thank belongs to my colleagues Dr. Dipak, Sakshi, Madhusudhan, Anjali, Harvijay, Siddhant, Shivli, Pooja, Dr. Satya and Dr. Preeti, for your friendship, support and all the inspiring scientific discussions during all these years. I would like to express my gratitude to all the fellow members, which although being part of different research group have indirectly or directly contributed their bit during my work. And a special thanks belongs to my dearest friend Jyoti

Singh Tomar, for always being there during all my ups and downs, for bearing my mood swings and tantrums, for her unconditional love and support, and for understanding and helping me throughout my this journey. In no words can I express my gratitude and the feeling that I am blessed, for her being an inseparable part of my life. I want to thank my parents and my sister for all the support now and before, and the very final thanks goes to my whimsical brother. Apologies if I may have forgotten to mention some names. Thank you for your love!

TABLE OF CONTENTS

LIST OF PUBLICATIONS

Tomar, Shailly, Manju Narwal, Etti Harms, Janet L. Smith, and Richard J. Kuhn. "Heterologous production, purification and characterization of enzymatically active Sindbis virus nonstructural protein nsP1." *Protein expression and purification* 79, no. 2 (2011): 277-284.

Pratap, Shivendra, Manju Narwal, Aditya Dev, Sonali Dhindwal, Shailly Tomar, and Pravindra Kumar. "Crimean-Congo Hemorrhagic Fever Virus: Strategies to Combat with an Emerging Threat to Human." *Current Bioinformatics* 7, no. 4 (2012): 467-477.

Development of colorimetric assay based on B-galactosidase for screening inhibitory compounds against alphaviral cysteine protease *(Manuscript in preparation)*

Crystal structure of Chikungunya nsP2 protease and biochemical characterization of its protease activity. *(Manuscript in preparation)*

WORKSHOPS/CONFERENCE ATTENDED AND POSTER PRESENTATION

Manju Narwal and Shailly Tomar . Studies of Sindbis virus replicase protein nsP1 7th Asian Biophysics Association Symposium and Annual Meeting of the Indian Biophysical Society (New Delhi),January 30-February 2, 2011.

Manju Narwal and Shailly Tomar . Structural insights into Alphavirus replicase protein nsP1. INDO-US Workshop/Symposium , Modern Trends in Macromolecular Structure (at IIT Bombay), February 21st-24th 2011.

Manju Narwal, Shivendra Pratap, Harvijay Singh, Pravindra Kumar and Shailly Tomar. Deciphering the structural features of the C-terminal domain of nsP2 replication protein. 42nd National Seminar on Crystallography (42nd-NSC) and International Workshop on Application of X-ray Diffraction in Drug Discovery (At JNU, NEW DELHI), November 21-23, 2013.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

The following list gives account of various abbreviations and acronyms used throughout the thesis

3D	3-Dimensional
Ala	Alanine
Arg	Arginine
Asn	Asparagine
Asp	Aspartic Acid
AURAV	Aura virus
CP	Capsid protein
CHIKV	Chikungunya virus
CSE	Conserved sequence element
Da	Dalton
DTT	Dithiothreitol
ddH20	Double distilled water
EEEV	Eastern equine encephalitis virus
ER	Endoplasmic reticulum
ELISA	Enzyme linked immune sorbent assay
Gln	Glutamine
Glu	Glutamic Acid
Gly	Glycine
Hepes acid	4-(2-hydroxyethyl)-l-piperazineethanesulfonic
His	Histidine
Leu	Leucine
Lys	Lysine
MES	2-(N-morpholino)ethanesulfonic acid

Met	Methionine
µ	Micron
NaCl	Sodium chloride
NTR's	Non-translated regions
NLS	Nuclear Localisation Signal
O.D. 600	Optical density at 600 nm
pI	Isoelectric point
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEG	Poly (ethylene glycol)
Phe	Phenylalanine
Pro	Proline
psi	Pounds per square inch
RT-PCR reaction	Reverse transciptase polymerase chain
r.m.s.d.	Root mean square deviation
RRV	Ross river virus
SAM	S-adenosyl L-methionine
SFV	Semliki forest virus
SINV	Sindbis Virus
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
Ser	Serine
Thr	Threonine
TCEP	tris(2-carboxyethyl)phosphine
Tris	tris(hydroxymethyl)aminomethane
Trp	Tryptophan

Tyr	Tyrosine
Val	Valine
UTR	Un-translated region
VEEV	Venezualan equine encephalitis virus
WEEV	Western equine encephalitis virus

Chapter I:

Introduction to alphaviruses

1. Introduction

1.1 Positive sense RNA viruses

Viruses are small infectious agents capable of multiplication only inside living host cells. Omnipresent as they may be called because of their presence in the nature, they have the capability of targeting organisms from all kingdoms of life. On the basis of the genetic matter, viruses are classified majorly into six different groups. The genetic matter could be DNA or RNA, and it could be either single or double stranded. Positive strand RNA viruses form the largest group of this classification. In these viruses the genetic matter i.e RNA is of the same sense as that of the host cell mRNA and could be directly translated to viral proteins by using host cell machinery [93]. Several human pathogens such as severe acute respiratory syndrome coronavirus (SARS-CoV), poliovirus, Hepatitis A, C and E viruses (HAV, HCV, and HEV), Rubella virus have positive-strand RNA as the genetic material. There is no DNA intermediate during the replication in these viruses; hence they maintain their genome purely in RNA form. The replication in positive sense RNA viruses progresses with the help of viral replication proteins, where an RNA dependent RNA polymerase (RdRp) unlike DNA polymerase helps in the synthesis of RNA from the RNA template. The replication process takes place in the cytoplasm of the host cell, mediated by the formation of membrane-bound replication complexes usually surrounded by the virus-modified host membranes. These replication complexes are the sites where all viral replication factors are concentrated for replication and for protection of viral RNA from host cell defense mechanism [76].

1.2 Alphavirus

Alphaviruses are arthropod borne viruses which belongs to *Togaviridae* family of virus classification, one which has positive sense RNA viruses as its members. Members of this family are distributed globally from North-South America to farthest south East Asia [110, 67]. Alphaviruses having a wide host range including humans, equines, birds, insects are capable of causing different levels of infections leading from mild rash, fever to severe encephalitis [30, 94]. Sindbis Virus, VEEV, Semliki Forest Virus and Ross River Virus are most extensively studied members of this genus. Chikungunya Virus, which was detected lately in 1952 even though of its clinical importance was not much of concern of research until the re-emergence of this virus in the year 2006 and that too at the global scale [89]. Even though the *genus* includes various pathogenic members, Sindbis Virus is

still being used as a prototype for understanding the alphaviral replication mechanism. Alphaviral genome consists of 11.7 kb of single stranded positive sense RNA with a methylguanylate cap at 5' terminus and a poly-adenylated tail at the 3'terminus. The 49S genomic matter produces mRNA which translates to give nsP1, nsP2, nsP3 and nsP4 during the viral infection inside the host cytoplasm. Initial translation of the genome gives two types of polyproteins i.e nsP123 and nsP1234. This happens because of an opal codon between nsP3 and nsP4 region. Mostly nsP123 is produced and it is a rare event when nsP1234 polyprotein synthesis takes place [94]. The polyprotein is sequentially cleaved into nsP4, nsP1, nsP3 and nsP2 and this sequential cleavage is found to be the regulator in the switch of minus and positive sense synthesis. During the Alphaviral replication, synthesis of negative sense RNA takes place from positive sense RNA which is executed by nsP123 and nsP4. The change in synthesis of RNA from minus strand to positive strand occurs when cleavage between nsp2/3 takes place [80]. RNA region near the 3' terminus which is approximate $1/3^{rd}$ region of the genomic RNA is translated to form structural polyproteins. The structural polyproteins are translated from 26S subgenomic RNA whose synthesis is also dependant on the regulated cleavage of non structural polyprotein.

1.2.1. Discovery and history of outbreaks

Discovery of alphaviruses dates back to the year 1930, when first member i.e Western equine encephalitis virus (WEEV) was isolated in United States [111]. However it was not until the WEEV epidemic in 1941, which incurred loss of around 3,00,000 equines and affected 4000 humans, which brought this family into the focus of researchers [111]. With this introduction, came the discovery of other alphavirus members. Although various outbreaks must have taken place previously, but this one with a major impact made researchers to study the pathophysiology of different members of this family. Followed by the discovery of WEEV, other member viruses were discovered; Eastern equine encephalitis virus (EEEV) in 1933 in eastern United States [26], Venezualan equine encephalitis virus (VEEV) in 1938 in Venezuela [5], Semliki forest virus (SFV) in 1942 in Uganda, Sindbis (SIN) was isolated in Cairo (Eqypt) in 1952 [97] and at the same time chikungunya virus (CHIKV) in *la reunion* [90]. O'nyong-nyong virus was emerged as a disease causing entity in the year 1959 in Uganda. A unusually distinct virus, O'nyong-nyong is more closely related to CHIKV than the other alphavirus members. Since 1959, this virus has caused at least three outbreaks in Africa affecting more than 2 million individuals [67]. Another member of this genus Mayaro virus was first isolated in 1954 and is enzootic in South American region [98]. Alphaviruses

have recently been identified to infect marine organisms like rainbow trout and Atlantic salmon pointing towards the increase in their host organisms and probably towards certain mutations enabling them to do so [54, 43].

1.2.2 Classification and geographical distribution

Alphaviruses are widely distributed group of viruses, classified under the *Togaviridae* family which belongs to Group IV of viruses in the Virus Classification. Alphavirus having around 30 members is one of the two genuses of the *togaviridae* family; other being Rubivirus with Rubella virus as the only member. Alphaviruses are distributed geographically to each and every continent except Antarctica, where although a new member affecting elephant seals has been discovered recently. The genus is divided into seven antigenic complexes, of which four medically important types are VEEV, EEEV, SFV and WEEV [94, 110, 67].

Alphaviruses are further grouped into New World and Old World viruses on the basis of their geographical distribution. New world viruses are widely distributed in Northern, Central as well as South America whereas Old world members cover the Europe, Africa, Asia and Australia continents. Old world viruses often causes rashes, fever and arthralgia with rare cases of mortality, whereas New world ones are known to cause fatal encephalitis. There is no clear evidence about the origin of the alphaviruses but on the basis of phylogeny studies it is assumed that they might have originated in New World countries and spread out to the Old world regions. Further evolution lead to the formation of two major groups Sindbis and Semliki forest. This divergence of New and Old world might have taken place some 2000-3000 years ago and some migratory birds or animals might have played role in this [30, 109].

WEEV and EEEV outbreaks are common in American continent especially United States region. They are transmitted by *Culiseta melanura* species of mosquitoes which help in the maintenance of enzootic and epizootic cycle of the viruses [117]. Generally WEEV and EEEV targets equines but sporadic cases of human infections are also reported. There are almost eight antigenic subtypes of VEEV having different virulence capabilities. Different species of *Aedes* and *Culex* helps in the transmission of VEEV into its different hosts. Birds are not a part of the infection cycle of the VEEV strains [87]. Strains with lesser virulence circulate in mosquito-rodent-mosquito enzootic cycle, whereas highly pathogenic strains able to cause encephalitis are able to infect humans crossing the epizootic cycle with the horses [112].

Table No. 1.1. Important alphaviruses causing infections in different vertebrate hosts (based on data from ref [81]).

Clinically Important Alphaviruses				
Virus	**Clinical syndrome**	**Vector**	**Hosts**	**Distribution**
Chikungunya (CHIKV)	Febrile illness, Rash , Arthralgia	Mosquito	Primates, Humans	Africa, India, SouthEast Asia
Eastern Equine Encepahlitis Virus (EEV)	Encephalitis	Mosquito	Birds	Americas
Mayaro Virus (MAY)	Febrile illness, Rash , Arthralgia	Mosquito	Primates, Humans	South America, Trinidad
O'nyong-nyong (ONN)	Febrile illness, Rash , Arthralgia	Mosquito	Primates	Africa
Ross River (RR)	Febrile illness, Rash , Arthralgia	Mosquito	Mammals, Humans	Australia, Pacific
Semliki Forest (SF)	Febrile illness, Rash , Encephalitis	Mosquito	Birds	Africa
Sindbis Virus (SIN)	Febrile illness, Rash , Arthralgia	Mosquito	Birds	Northern Europe, Africa, Australia, Asia
Venezualan Equine Encephalitis Virus (VEEV)	Febrile illness, Encephalitis	Mosquito	Rodents, Horses	Americas
Western Equine Encephalitis Virus (WEEV)	Encephalitis	Mosquito	Birds	North America

Chikunguya virus first observed in African regions during 1952 used to circulate in the primates and forest species of mosquitoes [36]. Then the transmission by *Aedes aegypti* led to the major epidemics in African and Southern Asian regions. Now the virus has become endemic to southern and southeastern Asia, although recent cases of CHIKV infection are reported in some Europen Countries also [90]. Mayaro Virus which is antigenically associated to CHIKV is native to Amazon River Basin. It circulates in primates and hematophagous mosquitoes but can infect humans on being in contact [98]. Ross River virus first discovered in 1959, from *Aedes vigilax* mosquitoes at the Ross river area in Australia is found to cause epidemic polyarthritis mostly in Australia and Pacific Ocean islands [116].

1.3 Alphaviral transmission cycle

Alphaviruses circulates in an enzootic and epizootic cycle in the environment. Arthropod vectors like mosquitoes acquire the infection after biting a viremic host, and after the replication of the virus in the vector during the incubation period transmits the infection to other hosts through salivary secretions [94]. Once inside the body of its host virus starts replication again which leads to viremia further causing illness. Humans, birds and other mammalian hosts generally acts as a principal hosts and are not part of the enzootic cycle which involves vectors which helps in the maintenance of the virus in the environment (Figure 1.1).

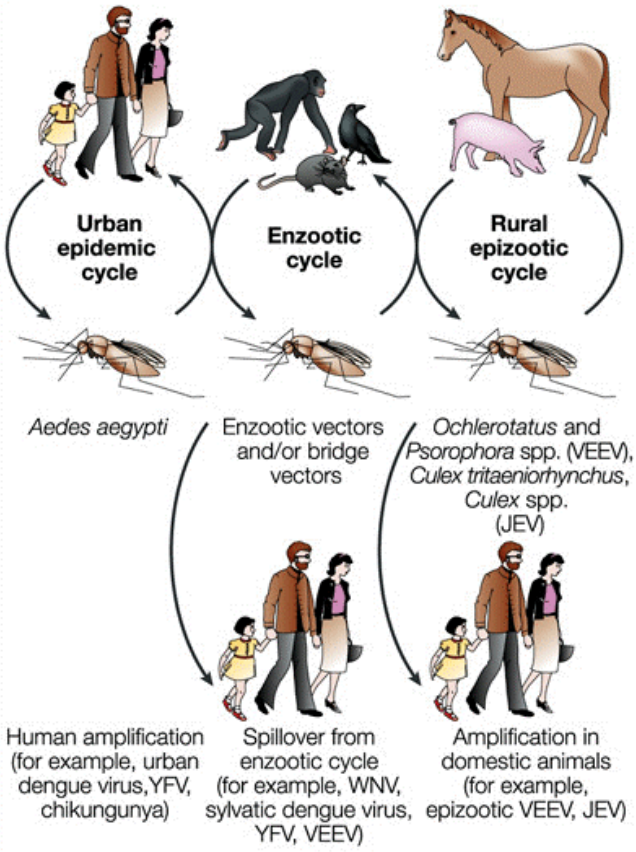


Figure 1.1 Life cycle of Alphavirus. Maintained in enzootic cycle by different species of mosquitoes and non-human primates. [21]

1.3.1. Mammalian hosts

Inside the principal host, first symptoms are visible only after the initial incubation period of 2-6 days, a period during which the virus tries to increase its copy numbers. There are further two phases of infection; acute phase and chronic phase. Lasting from few days to a couple of weeks the acute phase of infection includes symptoms like fever, chills, nausea, arthralgia (in

case of CHIKV, Mayaro and SIN) and in some cases maculopapular rashes (CHIKV infection). Clinical diagnosis of the infection in case of alphaviruses is difficult sometimes, since symptoms of most of the alphaviral diseases overlap (e.g it is difficult to establish infection for Ross river from rubella or enteriovirus infection) making it complicated to identify the disease in its initial stages. The fever in case of the infections does not stay for long but severe joint pain, headache and nausea can be suffered for 7-8 days. Cases where virus is capable of infecting the nervous system of the host (as in case of VEEV, EEEV and WEEV), the spread of infection is fast and in 5-10 days encephalitis is developed which may lead to death of the host [99, 119].

Alphaviruses affects different tissue cells of the host ranging from epithelial tissues to fibroblasts, macrophages and neurons (in case of encephalitis). The later stage of the infection i.e chronic phase lasts for a longer period of time. Symptoms are persistent and could be observed for a period of upto 2 years to life long also. Intermittent joint pain to myocarditis to meningoencephalitis and mild haemorrhage has been observed. Mortality rate however is observed to be low in case of viruses causing arthralgia or fever like symptoms but the post infection effects are more painful and long lasting [99, 119].

1.3.2. Mosquito vectors

Mosquito vectors which acquire the virus during the engorgement helps in the maintenance of alphaviruses in both enzootic and epizootic cycle. During feeding a host, mosquito ingest virus with the blood meal. This meal is transferred to mid-gut of the mosquito, now the place of virus initial replication. After two days of infection, virus reaches to other tissues and organs through the hemocoel. In the later stages, virus replication takes place in the salivary glands, after which the virus buds out from the infected cells and accumulates in the saliva. Once a vector gets infected with virus, it serves as a permanent host for life [15].

1.3.3. Disease treatment and control

Majority of the alphavirus members are pathogenic to humans and can easily be developed for use into bioterrorism but still no vaccine or antiviral therapy is available against them. Although trials are going on for vaccine development against CHIKV, but since majority of the vaccines developed are based on the use of live attenuated virus there are risk factors associated with its use. Vaccine development for Encephalitis virus like VEEV has been a

little success, but an attenuated strain of virus TC-83, providing protection to humans and equines have side effects and is non-effective in case of non-primates [2]. Treatment in case of members which causes arthralgia is generally symptomatic which includes using antihistamine drugs for treating rashes, non-salicylate analgesics and corticosteroids for relieving joint pain. Chloroquine treatment was used for treating Chikungunya infection; it although is effective but does have side effects and does not provide immunity for future infections [6, 8, 44].

1.4. The Biology of Alphaviruses

1.4.1. Genome organization and elements

The positive sense single stranded RNA genome in Alphaviruses is approximately 11.7 kb in length and accordingly produces proteins which help in the viral replication and infection mechanisms. The genome with the presence of 5' terminal methylguanylate cap, 5'and 3' untranslated regions (UTR's) and a 3' terminal polyadenylate tail has resemblance with cellular mRNA's. The overall organization of the Alphaviral genome seems to be conserved as the region coding for proteins which play role in the replication and replication complex formation is found at 5' proximal end and for the ones which play role in maintaining the virus structure are coded from 3' proximity. Proteins playing role in replication of the virus (generally quoted as Non-Structural proteins) are translated from 49S genomic RNA as a polyprotein [94]. Structural proteins are coded from a 26S sub-genomic RNA, which like genomic RNA is also capped at 5' end while polyadenylated at 3' terminus [94, 80].

Besides capping and adenylation, there are regions which are also conserved in the viral genomic RNA. Four highly conserved regions, known as *cis- acting* conserved sequence elements (CSEs) were identified at four different places in the genome. The regions where these CSEs are located are rendered important for playing role in the viral replication concluding that CSE's might be crucial for the replication of the virus. Located in the 5' UTR and in the early segment of 5' translated region are CSE1 and CSE2 respectively, which signifies the importance of the conservation of this region [94]. CSE3 is positioned in a domain which is at the juncture of the regions coding for structural and non structural proteins [48] and CSE4 is at UTR located at 3' terminus just before the polyadenylated tail region [66]. Mutational studies in these regions have shown to be affecting the viral growth, particularly replication [40, 48, 60, 83]. Many host and viral proteins have been studied to be binding to these CSE's, and these interactions have been assumed to be playing role in

promoting the replication [62, 63]. Also low rates of mutation have been observed in this region owing to the fact that any mutation in this region needs to be complemented with the mutation in the host proteins [61].

1.4.2. Structure of the virion

Alphaviruses are spherical shaped with around 60-70nm in diameter, enveloping the single stranded positive sense RNA genome. The surface envelope encapsidating the capsid protein is made up of E1 and E2 surface glycoproteins. The capsid protein with genomic RNA is arranged in T=4 icosahedral symmetry inside the envelope [94, 80].

The glycoproteins arranged as trimeric spikes, 80 in numbers, are embedded in the lipid bilayer derived from the host cell (Figure 1.2) [94]. These glycoproteins play role in the virus fusion with the host cell, penetration inside the cell and also in its exit from the cell via budding [35].

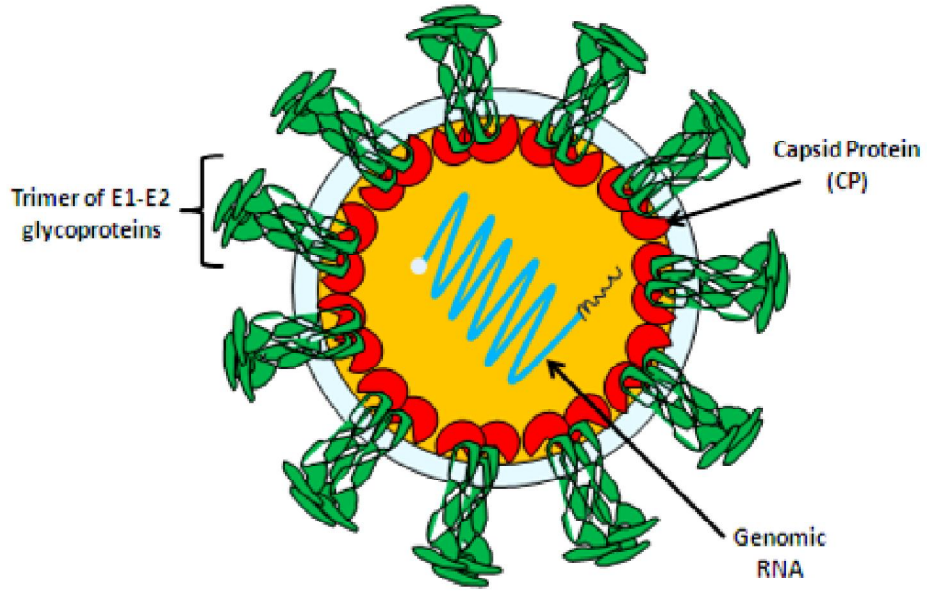


Figure 1.2 Structure of Alphavirus. Outer envelope made from the host cell membrane embedded with 80 E1-E2 glycoproteins trimeric spikes, encapsulating positive-sense single-stranded RNA. (Concept derived from http://viralzone.expasy.org/all_by_species/625.html)

1.4.3. Entry and disassembly

Entry of Alphavirus inside the host cell is considered to be an engagement between the glycoproteins and some host cell receptors. Although the exact mechanism is still elusive, but it is assumed by different research groups that either there is a fixed common receptor

like laminin (found on both eukaryotic or mosquitoes cell surface) with which the viral glycoproteins interact or could be a host of different cellular receptors with which the interaction might be taking place during the entry inside the host cell [38, 95, 105]. Based on the mutagenesis studies, E2 glycoprotein is understood to be contributing more than E1 glycoprotein in the interaction with the host cell receptor [17, 82]. Binding to the cell receptor induces conformational changes in E1 and E2 glycoproteins; evident with the fact that the transitional epitopes needed for the recognition by monoclonal bodies becomes available only after the binding of the glycoproteins with the receptors [57, 19]. Virus bound to the cell receptor is then endocytosed via clathrin dependant mechanism. With the maturation of this endocytic vesicle, pH drops inside and become acidic. This change in pH destabilizes the E1-E2 interaction and a fusion loop previously hidden and present at the end of the E1 glycoprotein is exposed. This loop then inserts into the endosomal membrane and as a result of this viral and endosomal membrane fuse together leading to the formation of a pore which paves the way to nucleocapsid (NC) into the cytoplasm. Inside the cytoplasm, the nucleocapsid assembly destabilizes separating genomic RNA from the capsid proteins to provide template for replication [94].

1.4.4. Replication of the alphaviral genome

The 49S genomic RNA once released into the cytoplasm is directly translated with the help of cellular translation machinery. The RNA is generally translated into nsP1234 or nsP123 polyproteins [51, 12]. The translation to the nsP1234 takes place when the polymerase read through the opal codon between the nsp3 and nsp4 protein [51, 12, 84]. The polyprotein is cleaved into different non structural proteins nsP1, nsP2, nsP3 and nsP4. The cleavage is the result of *cis* and *trans* activity of the papain like protease nsP2 [94].

There are different intermediate stages of the polyprotein cleavage and each cleavage regulates the different steps of the viral replication. The nsP1234 is first cleaved by *cis* activity of nsP2 leading to the formation of nsP123 and nsP4 RdRp (RNA dependant RNA polymerase) enzyme. In the early stage of the infection nsP123 is frequently formed and gets accumulated in large numbers. Then during the late stage of the infection, the nsP123 gets further cleaved into nsP1 and nsP23 by the *trans* activity of nsP2. One nsP123 molecule is cleaving the other nsP123 molecule in *trans* form. It is at this stage that finally nsP23 gets cleaved into nsP2 and nsP3 by the trans activity of the nsP2 protease [94].

Replication complex formation occurs on the vacuoles derived from the membranes of cytoplasmic oragenlles. Vacoules are identified to be the place where the positive sense

genomic RNA is reverse translated to form negative strand of the RNA with the help of the replication complex formed by the non structural proteins [3, 22].

There are different types of replication complexes which are formed during the replication. The nsP123 and nsP4 combine together to form a complex which helps in the synthesis of minus stranded RNA during the early stages of the infection. Since the concentration of nsP123 is high during early stages of infection, thus negative-strand synthesis is observed to be dominated over the positive-sense RNA synthesis [47, 84, 107]. As the infection proceeds a transient intermediate stage is observed to be formed by nsP1, nsP2, nsP3 and nsp4 which might be playing role in the negative and positive strand synthesis. This mature form of non structural proteins also leads to the synthesis of 26S sub-genomic RNA from the sub-genomic promoter on the genomic RNA [70]. During the later stages of the infection, nsP23 and nsP1 and nsP4 combine to form replication complex which leads to the synthesis of positive sense RNA (Figure 1.3) [94].

1.4.1.1. Non structural proteins

The non structural proteins forming the replication complex are formed as a polyprotein directly by the translation of the genomic RNA and mature into individual proteins by the step by step cleavage activity of nsP2 protease. All of the proteins play different but critical roles for the replication of the viral RNA. The nsP1 known as the capping enzyme posses the methyltransferase and guanyltransferase activity (Figure 1.4) needed for the capping of the genomic RNA. Unlike other capping enzymes, nsP1 is unique in terms of having both of the capping activities and different from other non structural proteins as it performs all of its functions only after attachment with the membrane [1, 58]. Amphipathic helix region of the nsP1 binds to the negatively charged phospholipids which facilitate this association, anchoring the replication complex on the membrane. Also, cysteine residues 418-420 were found to be palmitoylated which further strengthens this association of replication complex on the membrane [41, 1]. nsP1 has also been found to participate directly in initiation and elongation of minus-strand RNA synthesis of alphaviruses [31, 106] and this was suggested to require interaction with nsP4 [86]. The nsP2 is mainly divided into two domains having helicase and protease activity. Detailed discussion about our target is done separately [Section 2.1]. The nsP3 which is known as a phosphoprotein [50, 64] is not much studied in case of alphaviruses. It is basically divided into three domains; the first domain known as macro domain is highly conserved among alphaviruses, coronaviruses, Hepatisis E virus and

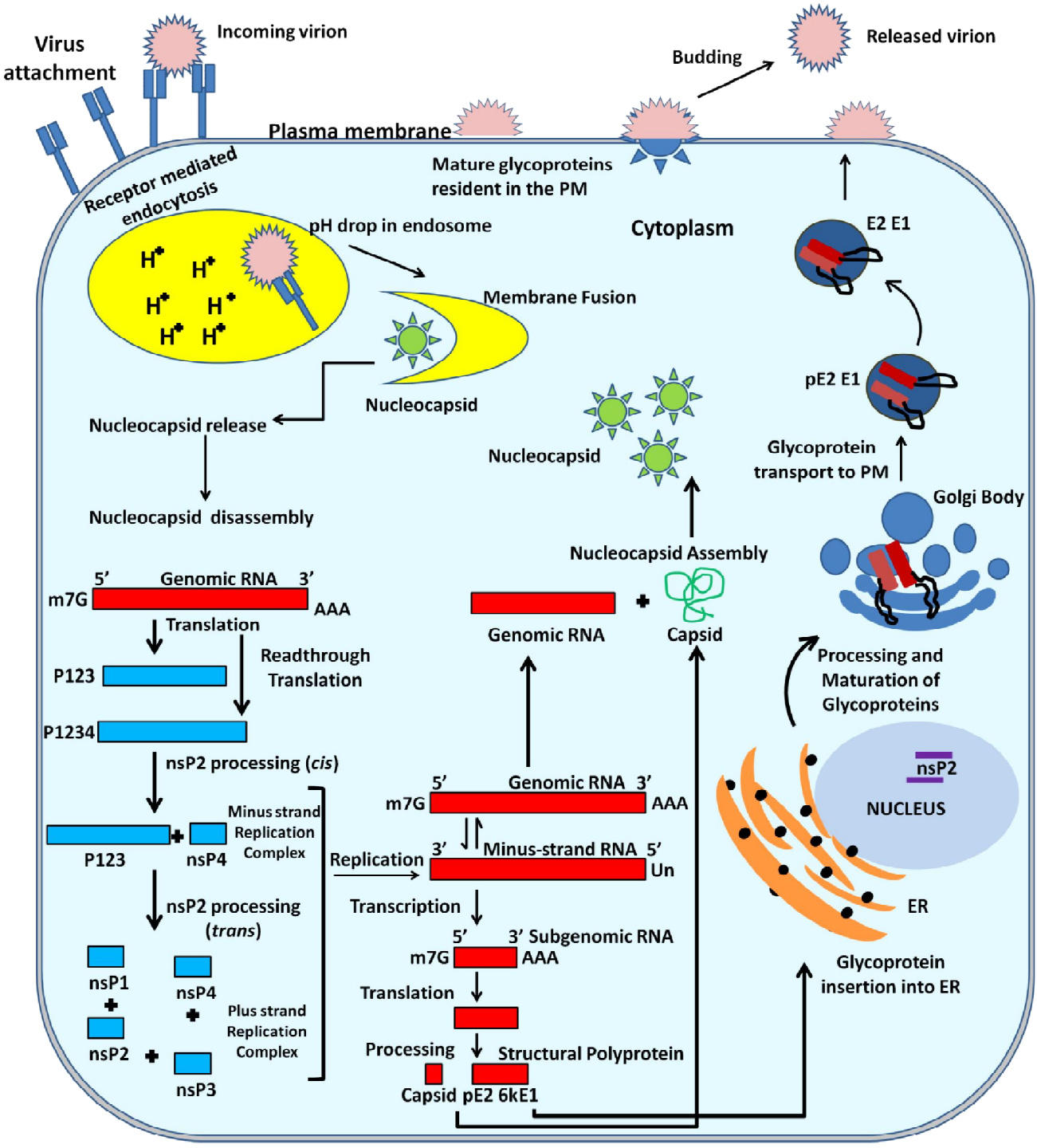
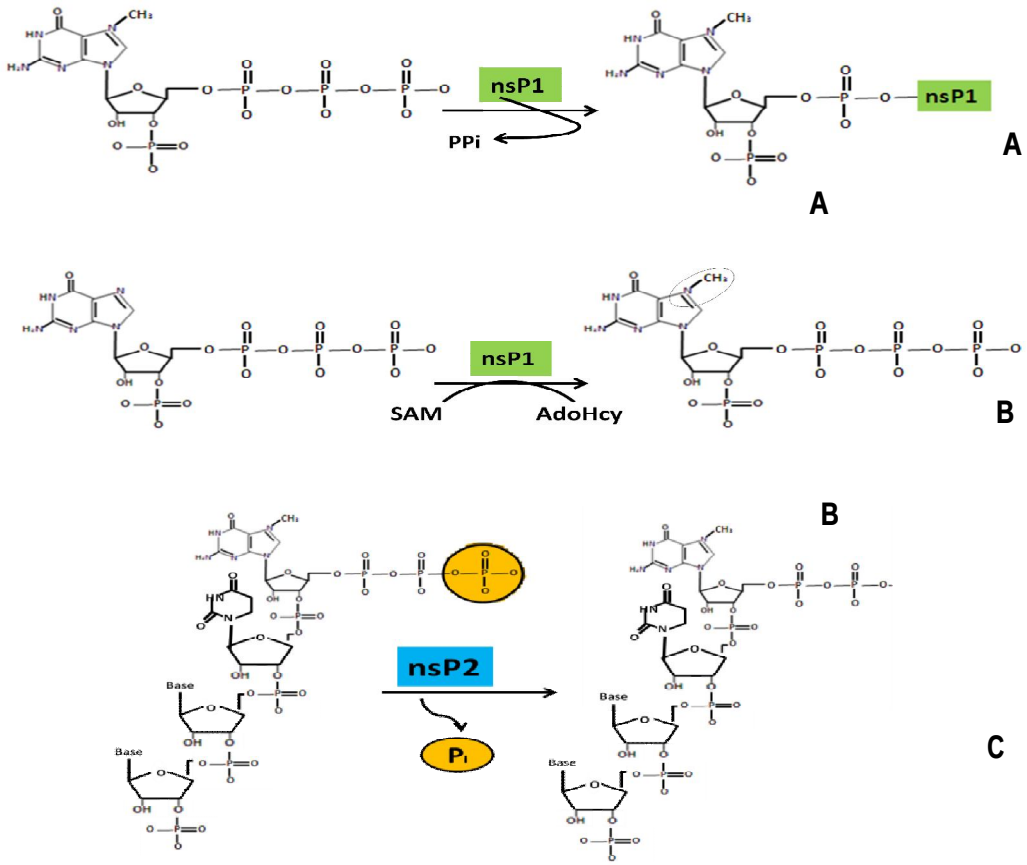


Figure 1.3 Alphavirus replication cycle. The RNA is translated into four non structural proteins which helps in the replication of the RNA in the cytoplasm. At the left bottom corner, RNA transcription and viral polyprotein processing are shown. Glycoproteins formed after the processing of viral structural polyprotein are further modified at endoplasmic reticulum (ER) and Golgi body as shown on the right side. Budding of the virus particles appears at the plasma membrane [Concept derived from 35].

Rubella virus, the second zinc binding domain (ZBD) is conserved only among alphaviruses and third C terminal hypervariable domain which is found to be important for virus replication [83]. This protein has been shown to be a phosphoprotein and majority of the phosphorylated serine and threoine residues are located in the hypervariable region of nsP3 [102, 103]. Deletions in this region affect the rates of RNA synthesis and replication pointing towards its possible role in the virus multiplication [31, 45].

The nsP4 is a RNA dependant RNA polymerase which play role in the synthesis of different types of viral RNA's [4, 33, 22, 78]. The concentration of nsP4 remains very low throughout the infection process owing to low concentration of nsP1234 polyprotein, which itself is dependent on the read-through beyond the opal codon between nsP3/4 [49, 13]. Possible nsP1-nsP4 interaction was suggested based on the identification of suppressor mutations in nsP1 along with substitutions at the N terminus of nsP4 with non-aromatic amino acids Ala, Arg and Leu. The N terminus of the nsP4 bore a Tyr residue which helps in the recognition of promoter for minus strand synthesis. Mutation studies has shown that substituting this N terminal Tyr is most likely affecting its interaction with nsP1 domain at residues A348-T349 affecting thus inhibiting minus strand synthesis [86].



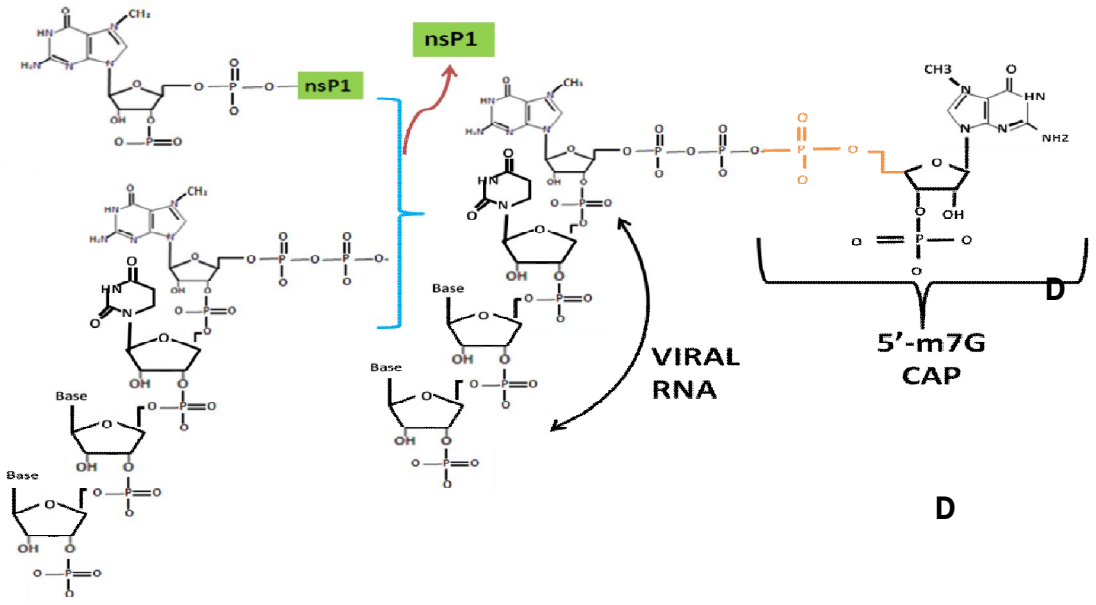


Figure 1.4 Capping of alphavirus RNA by nsP1 and nsP2 (A) Using SAM as a methyl donor, nsP1 methylates GTP at the 7 position. **(B)** then 7-methylGTP gets covalently attached to nsP1. **(C)** The 5'-triphosphate end of the viral RNA is then converted to 5'-diphosphate with the help of RNA 5'-triphosphatase activity of nsP2. **(D)** The 5'-end of the viral RNA is then recognized by 7-methyl-GMP-nsP1 and nsP1 finally catalyzes the transfer of 7-methylGMP to the 5'-end of the viral RNA.

1.4.1.2. Structural proteins, assembly and release

Structural proteins are expressed from 26S subgenomic RNA in a precursor form as Capsid/pE2/6k/E1 polyprotein. The polyprotein is processed by the host and viral proteases to develop individual structural proteins. Capsid protein (CP), a serine protease is the first one to act on the polyprotein chain [11, 88, 80]. Due to its *cis* autoproteolytic activity it cleaves and separates itself from the chain [114]. On cleavage, signal at the N terminus of the pE2 protein gets exposed which then directs the chain towards Endoplasmic reticulum [94, 80]

At ER, major modification and processing of the pE2 and E3 proteins takes place. In the polyprotein E1and E2 are glycoproteins and 6k acts as the signal sequence for E1 glycoprotein [37, 80]. The modification pathway of the glycoproteins takes course from ER to Golgi apparatus and terminates finally at Plasma membrane [80]. Host cells proteases like signalases and furin cleaves the polyprotein into pE2, 6k and E1 at ER, which is followed by

glycosylation at the N terminus of pE2 and E1 [80]. At golgi apparatus this oligosaccharide chain is further reduced as a part of modification and the pE2 is cleaved to finally separate E3 and E2 [55, 59, 77, 108]. This cleavage is important as uncleaved pE2 although would be able to perform budding but its capability to cause infection is lost [68, 74]. Cleaved CP protein recognizes a packaging signal located in the 5' region of the genome which signals it to encapsidate the genomic RNA leading to the formation of Nucleocapsid (NC) [114, 52, 115]. In the final stage of the infection, this NC interacts with the E2 at the plasma membrane [80]. E2 interacts with the CP through a hydrophobic pocket present on its surface [53]. This interaction leads to the attraction of more glycoproteins, resulting in the formation of glycoprotein coat around the NC which eventually buds out through the plasma membrane [9, 25]. Interactions between the E2 and E1glycoproteins are important for the assembly for these virus particles and then their budding from the cell [34, 92, 104].

1.5. Non structural protein 2 (nsP2)

Alphaviruses majorly consists of two different types of proteases which play role at the different steps of the virus replication: one serine protease, Capsid (CP) [10, 32] and another papain-like cysteine protease, nsP2 [100, 91]. The nsP2 protein is characterized mainly for the role it plays in the viral replication. The nsP2 has helicase [73, 27] and protease activities [91] and it also plays role in negative and positive sense RNA synthesis by being a part of nsP123 and nsP23 replication complexes [46, 101]. Not only in replication, there are other auxiliary roles observed to be associated with the nsP2.

1.5.1 Essential function in the replication complex

The nsP2 domain assertion is a cumbersome work since the two domains which are identified to be playing different roles somewhat overlaps each other. The N-terminal and C-terminal domains are assumed to be possessing helicase and protease activities respectively. The domain at the N terminus is a NTP binding domain having lysine dependant ATPase and GTPase activities. NTPase conserved motifs I (GSGKS) and II (DEAF) beginning from residues 189 and 250 respectively play role in NTP binding [79]. Sequence of these and other downstream conserved motifs share considerable homology with the motifs in superfamily 1 of helicases and also, nsP2 in different studies has been found to be able to bind the RNA molecules indicating towards its possible role in the unwinding of RNA during the replication [28, 42, 39]. Protease domain at the C-terminal of the nsP2 plays role

in the cleavage of the polyprotein nsP1234 into individual proteins. Sequential cleavage by the protease leads to the formation of different replication complexes at different stages of the replication [84, 85, 29]. First crystal structure of the protease domain was elucidated in VEEV, which gave insights about the different structural features of the C-terminal domain. From VEEV nsP2 protease structure, it is established that the C-terminal domain is further dissected into two domains: one N-terminal protease and another C-terminal methyltransferase [75]. The MTase-like domain has an overall significant tertiary structure similarity to Sadenosyl L-methionine-dependent MTase structures (e.g., Escherichia coli FtsJ and dengue virus NS5 [75, 79]. However, there was no significant similarity in the residues that correspond to the S-adenosyl L-methionine binding site between nsP2pro and FtsJ MTase, and structural alignment around the binding site was also poor [75]. Additionally, it has been suggested that the nsP2 MTase-like domain is not a functional MTase suggesting that this domain must have other non-enzymatic functions in viral RNA replication [75, 56].

The active site in the structure has the conformation of that of the papain proteases with cysteine and histidine at the critical positions [91]. Unlike the papain proteases it doesn't have asparagine as a part of catalytic triad. Mutagenesis studies have shown one tryptophan next to active site histidine, conserved among all of the alphaviruses to be the one playing role in the activity and is playing role in the formation of catalytic triad [91].

nsP2 protease cleaves at three different positions i.e nsP1/2, nsP2/3 and nsP3/4, where it recognizes specific sequences with a conserved glycine at P2 position (Figure 1.5, Figure 1.6) [120]. It also has different preference for all the sites, a fact observed by the kinetic studies of the protease with the peptides in case of SINV, VEEV and SFV [12, 118]. The change in template specificity of nsP2 in the replication may arise from differences in conformation of the proteins induced by cleavage. It was identified that the purified nsP2 protease is able to cleave nsP3/4 site peptide rapidly, nsP1/2 cleavage slowly and was not at all being able to act upon nsP2/3 site peptide. Recent research on nsP23zbd protein from SINV has hinted that nsP3 domain might be necessary for the cleavage at nsP2/3 position during virus replication [83]. This means that nsP2, also interacts with other non structural proteins and arrange in a conformation which acts as a platform for cleavage at different positions at the polyprotein. Mutagenesis studies beyond the defined protease domain demonstrate its role in the viral RNA synthesis. The nsP2 might be regulating the synthesis of negative-sense RNA, as the nsP2 mutants failed in the synthesis of positive-sense RNA from negative-sense RNA [118, 14]. The nsP2 also plays role in the synthesis of 26S subgenomic RNA by interacting with the subgenomic promoter present at the genomic RNA

[96]. These functions are performed by the different replication complexes which are formed by the cleavage of the nsP1234 polyprotein by the nsP2.

	nsP1 →				← nsP2								nsP2 →				← nsP3								nsP3 →				← nsP4			
	P4	*P3*	*P2*	*P1*	*P1′*	*P2′*	*P3′*	*P4′*	*P4*	*P3*	*P2*	*P1*	*P1′*	*P2′*	*P3′*	*P4′*					*P4*	*P3*	*P2*	*P1*	*P1′*	*P2′*	*P3′*	*P4′*				
SIN	D	I	G	A	A	L	V	E	G	V	G	A	A	P	S	Y	*/R	L	T	G	V	G	G	Y	I	F	S				
OCK	D	I	G	A	A	L	V	E	G	V	G	A	A	P	S	Y	*/R	L	T	G	V	G	G	Y	I	F	S				
AURA	D	A	G	A	A	L	V	E	G	S	G	A	A	P	S	Y	*/R	L	T	G	V	G	G	Y	I	F	S				
WHA	D	I	G	A	A	L	V	E	G	V	G	A	A	P	S	Y	*/R	L	T	G	V	G	G	Y	I	F	S				
VEEV	E	A	G	A	G	S	V	E	E	A	G	C	A	P	S	Y	*/R	R	F	D	A	G	A	Y	I	F	S				
EEEV	E	A	G	A	G	S	V	E	E	A	G	R	A	P	A	Y	*/R	R	Y	E	A	G	A	Y	I	F	S				
WEEV	E	A	G	A	G	S	V	E	E	A	G	R	A	P	A	Y	*/R	R	Y	E	A	G	A	Y	I	F	S				
ONN	R	A	G	A	G	I	V	E	R	A	G	C	A	P	S	Y	R	L	D	R	A	G	G	Y	I	F	S				
IO	R	A	G	A	G	I	V	E	R	A	G	C	A	P	S	Y	R	L	D	R	A	G	G	Y	I	F	S				
SFV	H	A	G	A	G	V	V	E	T	A	G	C	A	P	S	Y	R	L	G	R	A	G	A	Y	I	F	S				
RRV	R	A	G	A	G	V	V	E	T	A	G	C	A	P	S	Y	*/R	L	G	R	A	G	A	Y	I	F	S				
MID	R	A	G	A	G	V	V	N	T	A	G	C	A	P	S	Y	*/R	L	D	R	A	G	A	Y	I	F	S				
BFV	R	A	G	E	G	V	V	E	P	A	G	S	A	P	A	Y	*/R	L	G	R	A	G	G	Y	I	F	S				

Figure 1.5 Cleavage sites on nsP1234 polyprotein on which nsP2 protease acts. [120].

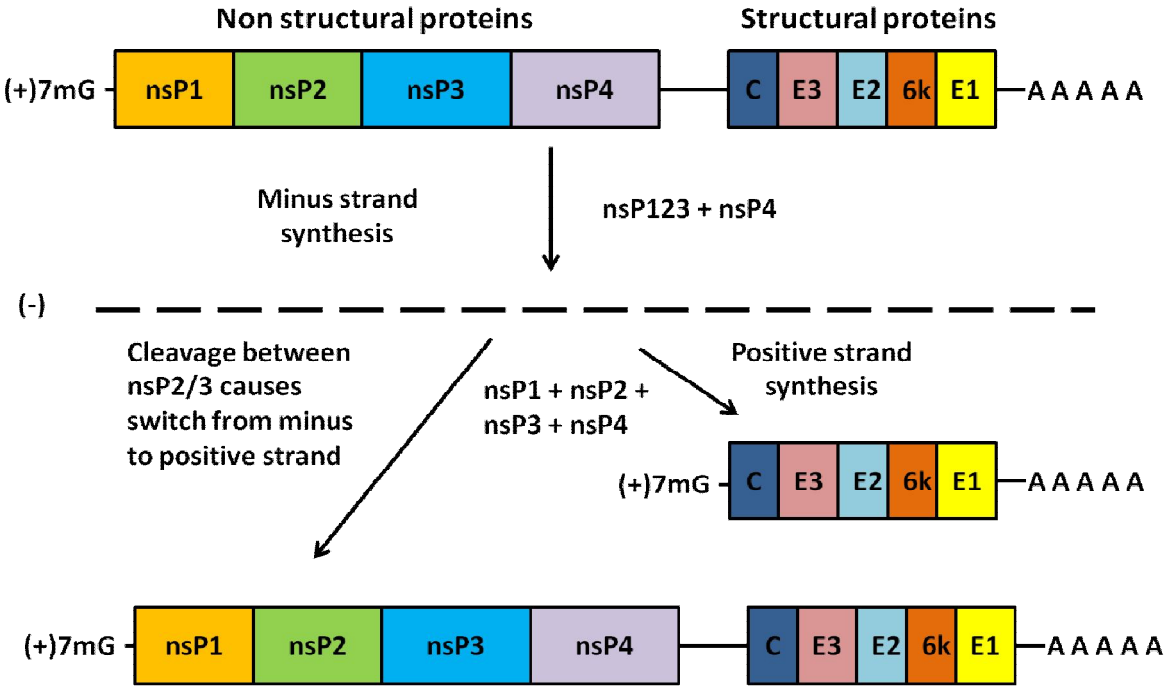


Figure 1.6 Cleavage of non structural polyprotein regulates the synthesis of non structural proteins nsP1, nsP2, nsP3 and nsP4. The cleavage is the check point of the positive, negative and 26S subgenomic RNA synthesis.

1.5.2. Evidence for auxillary functions

It is observed that along with playing role in the replication complex, there are more aspects of the viral infection in which nsP2 are imparting crucial contributions. Early indications for this came from the studies on SFV and SIN, in which nsP2 was located both in cytoplasm as well as in the nucleus of the infected cells, even after the establishment of the fact that viral replication takes place in cytoplasm [3, 71, 72]. Not only nsP2 was located in nucleus, but in case of SFV it was observed to be accumulated at the ribosomal assembly site i.e nucleolus [71]. Later on a pentapeptide sequence was discovered at the NLS region (Nuclear Localisation Signal) on SFV nsP2 consisting of positively charged arginine in the centre, which might be crucial for placing nsP2 in nucleus. Although it is unclear that what role nsP2 might be playing in the nucleus, but compromising the nsP2 NLS region affects the ability to reduce DNA synthesis by SFV nsP2 and also prevents the spread of the virus in the brain region [18].

There are other instances also where nsP2 has been observed to be affecting the outcome of infection. Alphaviruses are known to be causing lytic infections generally but when nsP2 has been modified by mutating some of its critical residues infections have been observed to be persistent and non-cytopathic showing the significance of nsP2 in the infection [16, 20, 65, 113]. nsP2 has also been observed to be cleaving the host proteins which are expressed as host antiviral response [7, 21]. Also, in old and new world alphaviruses difference is observed in the mechanism of host transcription shutoff. In old world viruses like SFVand SIN nsP2 shuts off the the host transcription but unable to do so in new world viruses like VEEV [23, 24]. In conclusion, nsP2 interacts with the host receptors and plays undefined role which is the major determinant of the infection virus is able to cause in the host cell.

1.6. Conclusion

Alphaviruses causes some of the most serious human diseases among all of the arboviruses. Non structural proteins of these viruses have overlapping roles in the viral RNA replication and also in interactions with the host factors. Our aim was to characterize this functional versatility of nonstructural proteins from SINV, CHIKV and AURA viruses. Over the years, a tremendous amount of information has been generated regarding the roles of alphavirus nonstructural proteins in replication. Three of the four nsPs have catalytic activities, and the interaction between the nsPs is dynamic and subject to temporal regulation driven by polyprotein processing. For understanding how the protease domain of nsP2 is performing its diverse functions and what possible structural changes it is undergoing for achieving these goals we targeted non structural protein nsP2 from three different alphaviruses.

Chapter 2 deals with the protease region of the multifunctional nonstructural protein, nsP2 of CHIKV. Conserved residues in this putative C-terminal domain were characterized for their roles in protease activity. Development of a colorimetry assay helped us in the understanding of the kinetics of the protease hinting about its preference order for cleavage at different recognition sites. Successful crystallization of the protein unravels the unique structural features of the cysteine protease.

Aura virus is not known to cause infection in humans and even though belonging to the old world of alphaviruses it is quite similar to the new world alphaviruses. It was thus interesting to target non structural protein from this virus for understanding if the virus despite being sequentially similar to the new world members has evolved or preserved the structural features also. Chapter 3 describes the purification and crystallization of aura nsP2 along with development of a simple β-galactosidase based protease assay.

And one important member of Alphavirus is Sindbis virus, which is mostly studied as model virus of the family. We targeted nsP2 from this member also so that we can establish a link between the information which we would have obtained from the structures of CHIKV and AURA nsP2's. However SIN nsP2 was quite unstable protein but with our efforts we were able to get a pure, stable protein which we have crystallized; described in chapter 4.

Chapter 5 describes comparison drawn among the available nsP2 structure information from CHIKV, VEEV and SIN viruses. Different flexible regions among the nsP2 protease structure

were observed and marked in case of CHIKV. Variations in conformations of active site dyad residues were observed.

1.7 Bibliography

1. Ahola, Tero, Anja Lampio, Petri Auvinen, and Leevi Kääriäinen. "Semliki Forest virus mRNA capping enzyme requires association with anionic membrane phospholipids for activity." The EMBO Journal 18, no. 11 (1999): 3164-3172.
2. Barrett, Alan DT, and Lawrence R. Stanberry. Vaccines for biodefense and emerging and neglected diseases. Academic Press, 2009.
3. Barton, DAVID J., STANLEY G. Sawicki, and DOROTHEA L. Sawicki. "Solubilization and immunoprecipitation of alphavirus replication complexes." Journal of virology 65, no. 3 (1991): 1496-1506.
4. Barton, DAVID J., S. G. Sawicki, and D. L. Sawicki. "Demonstration in vitro of temperature-sensitive elongation of RNA in Sindbis virus mutant ts6." Journal of virology 62, no. 10 (1988): 3597-3602.
5. Beck, C. E., and Ralph WG Wyckoff. "Venezuelan equine encephalomyelitis." American Association for the Advancement of Science, 1938.
6. Bettadapura, Jayaram, Lara J. Herrero, Adam Taylor, and Suresh Mahalingam. "Approaches to the treatment of disease induced by chikungunya virus." The Indian journal of medical research 138, no. 5 (2013): 762.
7. Breakwell, Lucy, Pia Dosenovic, Gunilla B. Karlsson Hedestam, Mauro D'Amato, Peter Liljeström, John Fazakerley, and Gerald M. McInerney. "Semliki Forest virus nonstructural protein 2 is involved in suppression of the type I interferon response." Journal of virology 81, no. 16 (2007): 8677-8684.
8. Brighton, S. W. "Chloroquine phosphate treatment of chronic Chikungunya arthritis." S Afr Med j 66, no. 6 (1984): 217-8.
9. Cadd, Tamarra L., Ulrica Skoging, and Peter Liljeström. "Budding of enveloped viruses from the plasma membrane." Bioessays 19, no. 11 (1997): 993-1000.

10. Choi, Hok-Kin, Liang Tong, Wladek Minor, Philippe Dumas, Ulrike Boege, Michael G. Rossmann, and Gerd Wengler. "Structure of Sindbis virus core protein reveals a chymotrypsin-like serine proteinase and the organization of the virion." (1991): 37-43.
11. Choi, Hok-Kin, Sukyeong Lee, Yan-Ping Zhang, Bonnie R. McKinney, Gerd Wengler, Michael G. Rossmann, and Richard J. Kuhn. "Structural analysis of Sindbis virus capsid mutants involving assembly and catalysis." Journal of molecular biology 262, no. 2 (1996): 151-167.
12. de Groot, Raoul J., W. Reef Hardy, Yukio Shirako, and James H. Strauss. "Cleavage-site preferences of Sindbis virus polyproteins containing the non-structural proteinase. Evidence for temporal regulation of polyprotein processing in vivo." The EMBO journal 9, no. 8 (1990): 2631.
13. de Groot, Raoul J., Tillmann Rümenapf, Richard J. Kuhn, Ellen G. Strauss, and James H. Strauss. "Sindbis virus RNA polymerase is degraded by the N-end rule pathway." Proceedings of the National Academy of Sciences 88, no. 20 (1991): 8967-8971.
14. De, Indra, Stanley G. Sawicki, and Dorothea L. Sawicki. "Sindbis virus RNA-negative mutants that fail to convert from minus-strand to plus-strand synthesis: role of the nsP2 protein." Journal of virology 70, no. 5 (1996): 2706-2719.
15. Domingo, Esteban, and John J. Holland. The origin and evolution of viruses. John Wiley & Sons, Ltd, 2008.
16. DRYGA, SERGEY A., OLGA A. DRYGA, and SONDRA SCHLESINGER. "Identification of mutations in a Sindbis virus variant able to establish persistent infection in BHK cells: the importance of a mutation in the nsP2 gene." Virology 228, no. 1 (1997): 74-83.
17. Dubuisson, J. E. A. N., and CHARLES M. Rice. "Sindbis virus attachment: isolation and characterization of mutants with impaired binding to vertebrate cells." Journal of virology 67, no. 6 (1993): 3363-3374.
18. Fazakerley, John K., Amanda Boyd, Marja L. Mikkola, and Leevi Kääriäinen. "A single amino acid change in the nuclear localization sequence of the nsP2 protein affects the neurovirulence of Semliki Forest virus." Journal of virology 76, no. 1 (2002): 392-396.
19. Flynn, D. C., W. J. Meyer, J. M. Mackenzie, and R. E. Johnston. "A conformational change in Sindbis virus glycoproteins E1 and E2 is detected at the plasma membrane as a consequence of early virus-cell interaction." Journal of virology 64, no. 8 (1990): 3643-3653.

20. Frolov, Ilya, Eugene Agapov, Thomas A. Hoffman, Béla M. Prágai, Mara Lippa, Sondra Schlesinger, and Charles M. Rice. "Selection of RNA replicons capable of persistent noncytopathic replication in mammalian cells." Journal of virology 73, no. 5 (1999): 3854-3865.
21. Frolova, Elena I., Rafik Z. Fayzulin, Susan H. Cook, Diane E. Griffin, Charles M. Rice, and Ilya Frolov. "Roles of nonstructural protein nsP2 and alpha/beta interferons in determining the outcome of Sindbis virus infection." Journal of virology 76, no. 22 (2002): 11254-11264.
22. Froshauer, Susan, Jiirgen Kartenbeck, and Ari Helenius. "Alphavirus RNA replicase is located on the cytoplasmic surface of endosomes and lysosomes." The Journal of cell biology 107, no. 6 (1988): 2075-2086.
23. Garmashova, Natalia, Rodion Gorchakov, Elena Frolova, and Ilya Frolov. "Sindbis virus nonstructural protein nsP2 is cytotoxic and inhibits cellular transcription." Journal of virology 80, no. 12 (2006): 5686-5696.
24. Garmashova, Natalia, Rodion Gorchakov, Eugenia Volkova, Slobodan Paessler, Elena Frolova, and Ilya Frolov. "The Old World and New World alphaviruses use different virus-specific proteins for induction of transcriptional shutoff." Journal of virology 81, no. 5 (2007): 2472-2484.
25. Garoff, Henrik, Roger Hewson, and Dirk-Jan E. Opstelten. "Virus maturation by budding." Microbiology and Molecular Biology Reviews 62, no. 4 (1998): 1171-1190.
26. Giltner, L. T., and M. S. Shahan. "The 1933 outbreak of infectious equine encephalomyelitis in the eastern states." North Am Vet 14 (1933): 25.
27. de Cedrón, Marta Gomez, Neda Ehsani, Marja L. Mikkola, Juan Antonio García, and Leevi Kääriäinen. "RNA helicase activity of Semliki Forest virus replicase protein NSP2." FEBS letters 448, no. 1 (1999): 19-22.
28. Gorbalenya, Alexander E., Vladimir M. Blinov, Alexei P. Donchenko, and Eugene V. Koonin. "An NTP-binding motif is the most conserved sequence in a highly diverged monophyletic group of proteins involved in positive strand RNA viral replication." Journal of molecular evolution 28, no. 3 (1989): 256-268.
29. Gorchakov, Rodion, Elena Frolova, Stanley Sawicki, Svetlana Atasheva, Dorothea Sawicki, and Ilya Frolov. "A new role for ns polyprotein cleavage in Sindbis virus replication." Journal of virology 82, no. 13 (2008): 6218-6231.

30. Griffin, D. E. Fields virology (Fourth ed. Vol. 1). Philadelphia: Lippincott Williams & Wilkins. (2001):917-962.
31. Hahn, Young S., Arash Grakoui, Charles M. Rice, Ellen G. Strauss, and James H. Strauss. "Mapping of RNA-temperature-sensitive mutants of Sindbis virus: complementation group F mutants have lesions in nsP4." Journal of virology 63, no. 3 (1989): 1194-1202.
32. Hahn, Chang S., and James H. Strauss. "Site-directed mutagenesis of the proposed catalytic amino acids of the Sindbis virus capsid protein autoprotease." Journal of virology 64, no. 6 (1990): 3069-3073.
33. Harrington, Patrick R., Boyd Yount, Robert E. Johnston, Nancy Davis, Christine Moe, and Ralph S. Baric. "Systemic, mucosal, and heterotypic immune induction in mice inoculated with Venezuelan equine encephalitis replicons expressing Norwalk virus-like particles." Journal of virology 76, no. 2 (2002): 730-742.
34. Hernandez, Raquel, Heuiran Lee, Christine Nelson, and Dennis T. Brown. "A single deletion in the membrane-proximal region of the Sindbis virus glycoprotein E2 endodomain blocks virus assembly." Journal of virology 74, no. 9 (2000): 4220-4228.
35. Jose, Joyce, Jonathan E. Snyder, and Richard J. Kuhn. "A structural and functional perspective of alphavirus replication and assembly." Future microbiology 4, no. 7 (2009): 837-856.
36. Jupp, P. G., and B. M. McIntosh. "Chikungunya virus disease." The arboviruses: epidemiology and ecology 2 (1988): 137-57.
37. Keegstra, K. E. N. N. E. T. H., B. A. R. T. H. O. L. O. M. E. W. Sefton, and D. A. V. I. D. Burke. "Sindbis virus glycoproteins: effect of the host cell on the oligosaccharides." Journal of virology 16, no. 3 (1975): 613-620.
38. Klimstra, William B., Kate D. Ryman, and Robert E. Johnston. "Adaptation of Sindbis virus to BHK cells selects for use of heparan sulfate as an attachment receptor." Journal of virology 72, no. 9 (1998): 7357-7366.
39. Koonin, Eugene V., Alexander E. Gorbalenya, Michael A. Purdy, Mikhail N. Rozanov, Gregory R. Reyes, and Daniel W. Bradley. "Computer-assisted assignment of functional domains in the nonstructural polyprotein of hepatitis E virus: delineation of an additional group of positive-strand RNA plant and animal viruses." Proceedings of the National Academy of Sciences 89, no. 17 (1992): 8259-8263.

40. Kuhn, Richard J., Zhang Hong, and James H. Strauss. "Mutagenesis of the 3'nontranslated region of Sindbis virus RNA." Journal of virology 64, no. 4 (1990): 1465-1476.
41. Laakkonen, Pirjo, Tero Ahola, and Leevi Kääriäinen. "The effects of palmitoylation on membrane association of Semliki Forest virus RNA capping enzyme." Journal of Biological Chemistry 271, no. 45 (1996): 28567-28571.
42. Lain, S., J. L. Riechmann, and J. A. Garcia. "RNA helicase: a novel activity associated with a protein encoded by a positive strand RNA virus." Nucleic acids research 18, no. 23 (1990): 7003-7006.
43. La Linn, May, Joy Gardner, David Warrilow, Grant A. Darnell, Clive R. McMahon, Ian Field, Alex D. Hyatt, Robert W. Slade, and Andreas Suhrbier. "Arbovirus of marine mammals: a new alphavirus isolated from the elephant seal louse, Lepidophthirus macrorhini." Journal of Virology 75, no. 9 (2001): 4103-4109.
44. Lamballerie, Xavier De, Véronique Boisson, Jean-Charles Reynier, Sébastien Enault, Rémi N. Charrel, Antoine Flahault, Pierre Roques, and Roger Le Grand. "On chikungunya acute infection and chloroquine treatment." Vector-Borne and Zoonotic Diseases 8, no. 6 (2008): 837-840.
45. LaStarza, Mark W., Julie A. Lemm, and Charles M. Rice. "Genetic analysis of the nsP3 region of Sindbis virus: evidence for roles in minus-strand and subgenomic RNA synthesis." Journal of virology 68, no. 9 (1994): 5781-5791.
46. Lemm, Julie A., T. Rümenapf, Ellen G. Strauss, James H. Strauss, and C. M. Rice. "Polypeptide requirements for assembly of functional Sindbis virus replication complexes: a model for the temporal regulation of minus-and plus-strand RNA synthesis." The EMBO journal 13, no. 12 (1994): 2925.
47. Lemm, Julie A., Anders Bergqvist, Carol M. Read, and Charles M. Rice. "Template-dependent initiation of Sindbis virus RNA replication in vitro." Journal of virology 72, no. 8 (1998): 6546-6553.
48. Levis, R., S. Schlesinger, and H. V. Huang. "Promoter for Sindbis virus RNA-dependent subgenomic RNA transcription." Journal of virology 64, no. 4 (1990): 1726-1733.
49. Li, G. P., and CHARLES M. Rice. "Mutagenesis of the in-frame opal termination codon preceding nsP4 of Sindbis virus: studies of translational readthrough and its effect on virus replication." Journal of virology 63, no. 3 (1989): 1326-1337.

50. Li, Guangpu, Mark W. La Starza, W. Reef Hardy, James H. Strauss, and Charles M. Rice. "Phosphorylation of Sindbis virus nsP3 in vivo and in vitro."Virology 179, no. 1 (1990): 416-427.
51. Li, G. U. A. N. G. P. U., and CHARLES M. Rice. "The signal for translational readthrough of a UGA codon in Sindbis virus RNA involves a single cytidine residue immediately downstream of the termination codon." Journal of virology67, no. 8 (1993): 5062-5067.
52. LINGER, BENJAMIN R., LYUDMYLA KUNOVSKA, RICHARD J. KUHN, and BARBARA L. GOLDEN. "Sindbis virus nucleocapsid assembly: RNA folding promotes capsid protein dimerization." Rna 10, no. 1 (2004): 128-138.
53. Lopez, Susana, Jian-Sheng Yao, Richard J. Kuhn, Ellen G. Strauss, and James H. Strauss. "Nucleocapsid-glycoprotein interactions required for assembly of alphaviruses." Journal of virology 68, no. 3 (1994): 1316-1323.
54. McLoughlin, M. F., and D. A. Graham. "Alphavirus infections in salmonids–a review." Journal of fish diseases 30, no. 9 (2007): 511-531.
55. Mayne, Jeffrey T., Charles M. Rice, Ellen G. Strauss, Michael W. Hunkapiller, and James H. Strauss. "Biochemical studies of the maturation of the small Sindbis virus glycoprotein E3." Virology 134, no. 2 (1984): 338-357.
56. Mayuri Geders, Todd W., Janet L. Smith, and Richard J. Kuhn. "Role for conserved residues of Sindbis virus nonstructural protein 2 methyltransferase-like domain in regulation of minus-strand synthesis and development of cytopathic infection." Journal of virology 82, no. 15 (2008): 7284-7297.
57. Meyer, WILLIAM J., and ROBERT E. Johnston. "Structural rearrangement of infecting Sindbis virions at the cell surface: mapping of newly accessible epitopes." Journal of virology 67, no. 9 (1993): 5117-5125.
58. Mi, Sha, Russell Durbin, Henry V. Huang, Charles M. Rice, and Victor Stollar. "Association of the Sindbis virus RNA methyltransferase activity with the nonstructural protein nsP1." Virology 170, no. 2 (1989): 385-391.
59. Moehring, Joan M., N. M. Inocencio, B. J. Robertson, and T. J. Moehring. "Expression of mouse furin in a Chinese hamster cell resistant to Pseudomonas exotoxin A and viruses complements the genetic lesion." Journal of Biological Chemistry 268, no. 4 (1993): 2590-2594.

60. Niesters, H. G., and James H. Strauss. "Defined mutations in the 5'nontranslated sequence of Sindbis virus RNA." Journal of virology 64, no. 9 (1990): 4162-4168.
61. Niesters, H. G., and James H. Strauss. "Mutagenesis of the conserved 51-nucleotide region of Sindbis virus." Journal of virology 64, no. 4 (1990): 1639-1647.
62. Pardigon, Nathalie, Edith Lenches, and James H. Strauss. "Multiple binding sites for cellular proteins in the 3'end of Sindbis alphavirus minus-sense RNA." Journal of virology 67, no. 8 (1993): 5003-5011.
63. Pardigon, Nathalie, and James H. Strauss. "Mosquito homolog of the La autoantigen binds to Sindbis virus RNA." Journal of virology 70, no. 2 (1996): 1173-1181.
64. Pardigon, Nathalie, and James H. Strauss. "Mosquito homolog of the La autoantigen binds to Sindbis virus RNA." Journal of virology 70, no. 2 (1996): 1173-1181.
65. Perri, Silvia, David A. Driver, Jason P. Gardner, Scott Sherrill, Barbara A. Belli, Thomas W. Dubensky, and John M. Polo. "Replicon vectors derived from Sindbis virus and Semliki forest virus that establish persistent replication in host cells." Journal of virology 74, no. 20 (2000): 9802-9807.
66. Pfeffer, Martin, Richard M. Kinney, and Oskar-Rüger Kaaden. "The alphavirus 3′-nontranslated region: size heterogeneity and arrangement of repeated sequence elements." Virology 240, no. 1 (1998): 100-108.
67. Powers, Ann M., Aaron C. Brault, Yukio Shirako, Ellen G. Strauss, WenLi Kang, James H. Strauss, and Scott C. Weaver. "Evolutionary relationships and systematics of the alphaviruses." Journal of Virology 75, no. 21 (2001): 10118-10131.
68. Presley, J. F., J. M. Polo, R. E. Johnston, and D. T. Brown. "Proteolytic processing of the Sindbis virus membrane protein precursor PE2 is nonessential for growth in vertebrate cells but is required for efficient growth in invertebrate cells." Journal of virology 65, no. 4 (1991): 1905-1909.
69. Raju, R. A. M. A. S. W. A. M. Y., and HENRY V. Huang. "Analysis of Sindbis virus promoter recognition in vivo, using novel vectors with two subgenomic mRNA promoters." Journal of virology 65, no. 5 (1991): 2501-2510.
70. Rice, Charles M. "Examples of expression systems based on animal RNA viruses: alphaviruses and influenza virus." Current opinion in biotechnology 3, no. 5 (1992): 523-532.

71. Rikkonen, Marja, Johan Peranen, and Leevi Kaariainen. "Nuclear and nucleolar targeting signals of Semliki Forest virus nonstructural protein nsP2." Virology189, no. 2 (1992): 462-473.
72. Rikkonen, Marja, J. Peränen, and L. Kääriäinen. Nuclear targeting of Semliki Forest virus nsP2. Springer Vienna, 1994.
73. Rikkonen, Marja, J. Peränen, and L. Kääriäinen. "ATPase and GTPase activities associated with Semliki Forest virus nonstructural protein nsP2."Journal of virology 68, no. 9 (1994): 5804-5810.
74. Russell, DARCY L., JOEL M. Dalrymple, and R. E. Johnston. "Sindbis virus mutations which coordinately affect glycoprotein processing, penetration, and virulence in mice." Journal of virology 63, no. 4 (1989): 1619-1629.
75. Russo, Andrew T., Mark A. White, and Stanley J. Watowich. "The crystal structure of the Venezuelan equine encephalitis alphavirus nsP2 protease."Structure 14, no. 9 (2006): 1449-1458.
76. Salonen, A. N. N. E., T. E. R. O. Ahola, and L. E. E. V. I. Kääriäinen. "Viral RNA replication in association with cellular membranes." In Membrane trafficking in viral replication, pp. 139-173. Springer Berlin Heidelberg, 2005.
77. Sariola, Merja, Jaakko Saraste, and Esa Kuismanen. "Communication of post-Golgi elements with early endocytic pathway: regulation of endoproteolytic cleavage of Semliki Forest virus p62 precursor." Journal of Cell Science 108, no. 6 (1995): 2465-2475.
78. Sawicki, Dorothea L., David B. Barkhimer, Stanley G. Sawicki, Charles M. Rice, and Sondra Schlesinger. "Temperature sensitive shut-off of alphavirus minus strand RNA synthesis maps to a nonstructural protein, nsP4." Virology174, no. 1 (1990): 43-52.
79. Sawicki, Dorothea L., Silvia Perri, John M. Polo, and Stanley G. Sawicki. "Role for nsP2 proteins in the cessation of alphavirus minus-strand synthesis by host cells." Journal of virology 80, no. 1 (2006): 360-371.
80. Schlesinger, Sondra., and Milton J., Schlesinger. Fields virology (Fourth ed. Vol. 1). Philadelphia: Lippincott Williams & Wilkins. 895-916.
81. Schmaljohn, Alan L., and David McClain. "Alphaviruses (togaviridae) and flaviviruses (flaviviridae)." Medical Microbiology 54 (1996).

82. Schoepp, Randal J., and Robert E. Johnston. "Directed mutagenesis of a Sindbis virus pathogenesis site." Virology 193, no. 1 (1993): 149-159.
83. Shin, Gyehwa, Samantha A. Yost, Matthew T. Miller, Elizabeth J. Elrod, Arash Grakoui, and Joseph Marcotrigiano. "Structural and functional insights into alphavirus polyprotein processing and pathogenesis." Proceedings of the National Academy of Sciences 109, no. 41 (2012): 16534-16539.
84. Shirako, Yukio, and James H. Strauss. "Regulation of Sindbis virus RNA replication: uncleaved P123 and nsP4 function in minus-strand RNA synthesis, whereas cleaved products from P123 are required for efficient plus-strand RNA synthesis." Journal of Virology 68, no. 3 (1994): 1874-1885.
85. Shirako, Yukio, and James H. Strauss. "Cleavage between nsP1 and nsP2 initiates the processing pathway of Sindbis virus nonstructural polyprotein P123." Virology 177, no. 1 (1990): 54-64.
86. Shirako, Yukio, Ellen G. Strauss, and James H. Strauss. "Suppressor mutations that allow Sindbis virus RNA polymerase to function with nonaromatic amino acids at the N-terminus: evidence for interaction between nsP1 and nsP4 in minus-strand RNA synthesis." Virology 276, no. 1 (2000): 148-160.
87. Shope, ROBERT E. "Medical significance of togaviruses: an overview of diseases caused by togaviruses in man and in domestic and wild vertebrate animals." The togaviruses (1980): 47-82.
88. Skoging-Nyberg, Ulrica, and Peter Liljeström. "MXI motif of Semliki Forest virus capsid protein affects nucleocapsid assembly." Journal of virology 75, no. 10 (2001): 4625-4632.
89. Sourisseau, Marion, Clémentine Schilte, Nicoletta Casartelli, Céline Trouillet, Florence Guivel-Benhassine, Dominika Rudnicka, Nathalie Sol-Foulon et al. "Characterization of reemerging chikungunya virus." PLoS Pathog 3, no. 6 (2007): e89.
90. Staples, J. Erin, Robert F. Breiman, and Ann M. Powers. "Chikungunya fever: an epidemiological review of a re-emerging infectious disease." Clinical Infectious Diseases 49, no. 6 (2009): 942-948.
91. Strauss, Ellen G., Raoul J. De Groot, Randy Levinson, and James H. Strauss. "Identification of the active site residues in the nsP2 proteinase of Sindbis virus." Virology 191, no. 2 (1992): 932-940.

92. Strauss, Ellen G., Edith M. Lenches, and James H. Strauss. "Molecular genetic evidence that the hydrophobic anchors of glycoproteins E2 and E1 interact during assembly of alphaviruses." Journal of virology 76, no. 20 (2002): 10188-10194.
93. Strauss, Ellen G., and James H. Strauss. Viruses and human disease. Academic Press, 2007.
94. Strauss, James H., and Ellen G. Strauss. "The alphaviruses: gene expression, replication, and evolution." Microbiological reviews 58, no. 3 (1994): 491.
95. Strauss, J. H., K-S. Wang, A. L. Schmaljohn, R. J. Kuhn, and E. G. Strauss.Host-cell receptors for Sindbis virus. Springer Vienna, 1994.
96. Suopanki, Jaana, Dorothea L. Sawicki, and Stanley G. Sawicki. "Regulation of alphavirus 26S mRNA transcription by replicase component nsP2." Journal of general virology 79, no. 2 (1998): 309-319.
97. Taylor, R. M., H. S. Hurlbut, T. H. Work, J. R. Kingston, and T. E. Frothingham. "Sindbis virus: a newly recognized arthropod-transmitted virus." The American journal of tropical medicine and hygiene 4, no. 5 (1955): 844-862.
98. Tesh, Robert B., Douglas M. Watts, Kevin L. Russell, Chitra Damodaran, Carlos Calampa, Cesar Cabezas, Gladys Ramirez et al. "Mayaro virus disease: an emerging mosquito-borne zoonosis in tropical South America." Clinical infectious diseases 28, no. 1 (1999): 67-73.
99. Tyring, Stephen, Angela Yen Moore, and Omar Lupi, eds. Mucocutaneous Manifestations of Viral Diseases: An Illustrated Guide to Diagnosis and Management. CRC press, 2010.
100. Vasiljeva, Lidia, Leena Valmu, Leevi Kääriäinen, and Andres Merits. "Site-specific protease activity of the carboxyl-terminal domain of Semliki Forest virus replicase protein nsP2." Journal of Biological Chemistry 276, no. 33 (2001): 30786-30793.
101. Vasiljeva, Lidia, Andres Merits, Andrey Golubtsov, Valeria Sizemskaja, Leevi Kääriäinen, and Tero Ahola. "Regulation of the sequential processing of Semliki Forest virus replicase polyprotein." Journal of Biological Chemistry 278, no. 43 (2003): 41636-41645.
102. Vihinen, Helena, and Juhani Saarinen. "Phosphorylation site analysis of Semliki forest virus nonstructural protein 3." Journal of Biological Chemistry 275, no. 36 (2000): 27775-27783.

103. Vihinen, Helena, Tero Ahola, Minna Tuittila, Andres Merits, and Leevi Kääriäinen. "Elimination of phosphorylation sites of Semliki Forest virus replicase protein nsP3." Journal of Biological Chemistry 276, no. 8 (2001): 5745-5752.

104. Von Bonsdorff, C. H., and S. C. Harrison. "Hexagonal glycoprotein arrays from Sindbis virus membranes." Journal of virology 28, no. 2 (1978): 578-583.

105. Wang, KANG-SHENG, R. J. Kuhn, E. G. Strauss, S. U. S. A. N. Ou, and J. H. Strauss. "High-affinity laminin receptor is a receptor for Sindbis virus in mammalian cells." Journal of virology 66, no. 8 (1992): 4992-5001.

106. Wang, Y. F., STANLEY G. Sawicki, and DOROTHEA L. Sawicki. "Sindbis virus nsP1 functions in negative-strand RNA synthesis." Journal of virology 65, no. 2 (1991): 985-988..

107. Wang, Yun-Fen, Stanley G. Sawicki, and Dorothea L. Sawicki. "Alphavirus nsP3 functions to form replication complexes transcribing negative-strand RNA." Journal of virology 68, no. 10 (1994): 6466-6475.

108. Watson, D. G., J. M. Moehring, and T. J. Moehring. "A mutant CHO-K1 strain with resistance to Pseudomonas exotoxin A and alphaviruses fails to cleave Sindbis virus glycoprotein PE2." Journal of virology 65, no. 5 (1991): 2332-2339.

109. Weaver, S. C., A. Hagenbaugh, L. A. Bellew, S. V. Netesov, V. E. Volchkov, G. J. Chang, D. K. Clarke, L. Gousset, T. W. Scott, and D. W. Trent. "A comparison of the nucleotide sequences of eastern and western equine encephalomyelitis viruses with those of other alphaviruses and related RNA viruses." Virology 202, no. 2 (1994): 1083.

110. Weaver, Scott C., Wenli Kang, Yukio Shirako, Tillman Rumenapf, Ellen G. Strauss, and James H. Strauss. "Recombinational history and molecular evolution of western equine encephalomyelitis complex alphaviruses." Journal of virology 71, no. 1 (1997): 613-623.

111. Weaver, Scott C., Cristina Ferro, Roberto Barrera, Jorge Boshell, and Juan-Carlos Navarro. "Venezuelan equine encephalitis*." Annual Reviews in Entomology 49, no. 1 (2004): 141-174.

112. Weaver, Scott C., and Alan DT Barrett. "Transmission cycles, host range, evolution and emergence of arboviral disease." Nature Reviews Microbiology 2, no. 10 (2004): 789-801.

113. Weiss, B. A. R. B. A. R. A., R. I. C. H. A. R. D. Rosenthal, and S. O. N. D. R. A. Schlesinger. "Establishment and maintenance of persistent infection by Sindbis virus in BHK cells." Journal of virology 33, no. 1 (1980): 463-474.

114. Weiss, B. A. R. B. A. R. A., H. A. N. S. Nitschko, I. N. G. R. I. D. Ghattas, R. E. B. E. C. C. A. Wright, and S. O. N. D. R. A. Schlesinger. "Evidence for specificity in the encapsidation of Sindbis virus RNAs." Journal of virology 63, no. 12 (1989): 5310-5318.

115. Weiss, Barbara, Ute Geigenmüller-Gnirke, and Sondra Schlesinger. "Interactions between Sindbis virus RNAs and a 68 amino acid derivative of the viral capsid protein further defines the capsid binding site." Nucleic acids research 22, no. 5 (1994): 780-786.

116. Yu, Weiwei, Kerrie Mengersen, Pat Dale, John S. Mackenzie, Ghasem Sam Toloo, Xiaoyu Wang, and Shilu Tong. "Epidemiologic patterns of Ross River virus disease in Queensland, Australia, 2001–2011." The American journal of tropical medicine and hygiene 91, no. 1 (2014): 109-118.

117. Zacks, Michele A., and Slobodan Paessler. "Encephalitic alphaviruses."Veterinary microbiology 140, no. 3 (2010): 281-286.

118. Zhang, Di, József Tözsér, and David S. Waugh. "Molecular cloning, overproduction, purification and biochemical characterization of the p39 nsp2 protease domains encoded by three alphaviruses." Protein expression and purification 64, no. 1 (2009): 89-97.

119. Zuckerman, Arie J. Principles and practice of clinical virology. John Wiley & Sons, 2009.

120. ten Dam, Edwin, Michael Flint, and Martin D. Ryan. "Virus-encoded proteinases of the Togaviridae." Journal of general virology 80, no. 8 (1999): 1879-1888..

Chapter II

Chapter II:

Chikungunya nsP2 papain-like protease expression, purification, characterization and crystallization.

Chapter II

2.1 Introduction

Chikungunya virus is small approximately 70 nm size, positive-sense virus [19, 54] which belongs to the Old world group of the Alphaviruses. Chikungunya is a part of Semliki Forest SF group of the Old world of the Alphaviruses [27]. The Old world group members are generally associated with the polyarthritis and rash like symptoms, whereas encephalitis is termed to be the feature of the new world members [15, 37, 42]. Even though chikungunya belongs to the old group of viruses assuming it to be an arthritogenic alphavirus, but cases of hemorrhagic disease and meningoencephalitis has been observed in its case [13].

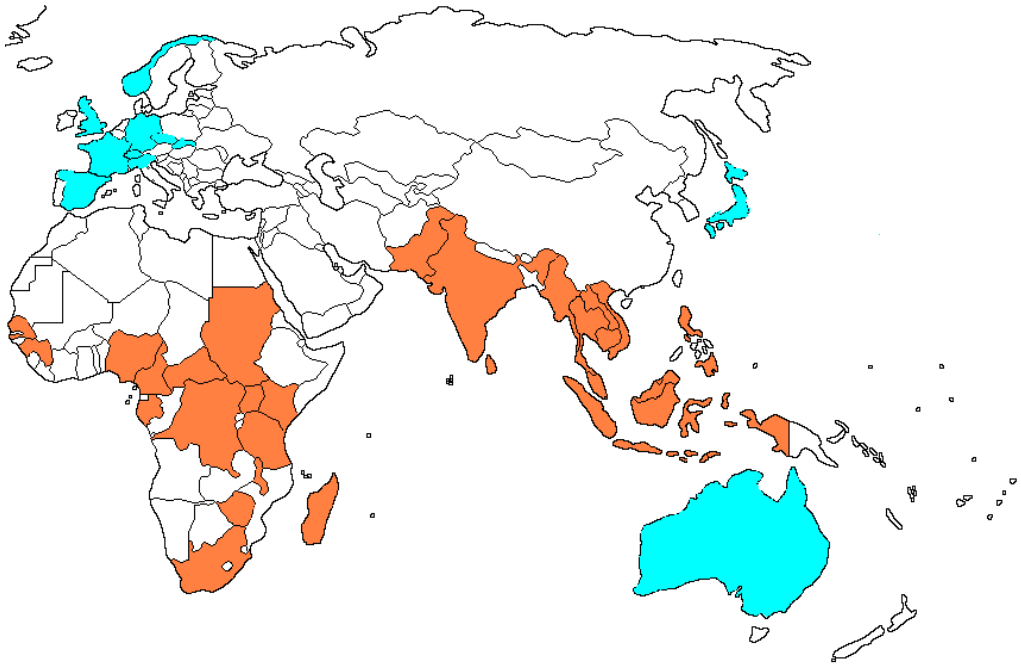


Figure 2.1 Distribution of chikungunya virus in 2010 (Old world countries). Both green and orange indicate countries where cases of chikungunya fever have been documented, and orange indicates countries where chikungunya virus has been endemic or epidemic [50].

Chikungunya is likely to be originated from the African region, where first case of the outbreak was reported in Makonde Plateau in 1952 [29, 48]. Subsequent outbreaks of chikungunya in Bangkok [36] and India were reported in 1960's. Outbreaks in different Indian cities like Calcutta, Maharashtra and Yellore were reported till the year 1973 [39, 53]. After that sporadic

cases of the chikungunya infections were observed for next 30 years, but no major outbreak took place. In the year 2004, two major outbreaks of chikungunya were reported in Kenya [43,13] which marked the beginning of a four year period during which the outbreak spread to different parts of India, numerous islands in Indian Ocean and some parts of Southeast Asia with the counting of number of affected countries to 40 [6, 51, 52].

Figure 2.1 Distribution of chikungunya virus in New world countries. 2015 CDC site.

The infection spread up to La Reunion island in the year 2006, which affected around 3,00,000 inhabitants of the island and killing around 250 people [3, 6]. Chikungunya is generally transmitted by the *Aedes aedypti* strains of the mosquito; however during the outbreak of the virus in the South Asian region *Aedes albopictus* species was found to be contributing in the spread of the virus [46, 63]. In chikungunya strain isolated from La Reunion, single point mutation was observed in the E1 glycoprotein which has resulted in the affinity of the virus towards other *Aedes* strains. This E1- A226V mutation has not only lead to the increase in the number of the vector hosts, but also due to this mutation there is increase in the infectivity by *Aedes albopictus* [46, 63, 59]. During this period the virus normally convicted of causing infections in the Old world regions has travelled to European [40, 47] and US regions also [43]. In 2007 September, aboriginal infection of the virus causes epidemic in Italy, which affected 200

people. After this, in the year 2008 chikungunya was listed in the US National Institute of Allergy and Infectious Diseases (NIAID) category C priority pathogen [43, 57].

Normally chikungunya infection has incubation period of 2-6 days followed by 4-7 days post infection period during which infection symptoms appear. Since most of the symptoms of chikungunya are similar to dengue infection, it sometimes leads to misdiagnosis of the infection eventually causing delay in the treatment of the diseased [17, 35, 36, 57]. Upon infection the virus presents itself in two phases; acute phase and chronic phase [9, 18]. The acute phase of the infection lasts from few days to couple of weeks with the characteristic features like chills, high fever, nausea, head ache, arthralgia and in some cases macupapular rashes [15, 42, 64]. The pain due to arthralgia is intense that it incapacitates its victim by making any kind of movement difficult [65, 33]. The virus targets different parts of the host by infecting the epithelial cells, fibroblasts and monocyte- derived macrophages [7].

Table 2.1 Disease symptoms of CHIKV and dengue infection; many features overlap in case of both the infections making one indistinguishable from the other [57].

Clinical features	Chikungunya	Dengue
Fever	+++	++
Arthalgias	+++	+/-
Lymphopenia	+++	++
Rash	++	+
Leukopenia	++	+++
Myalgias	+	++
Headache	+	++
Neutropenia	+	+++
Thrombocytopenia	+	+++
Bleeding dyscrasias	+/-	++
Shock	-	+/-

The chronic stage of arthalgia is experienced in the patients for a longer duration which could last from few months to years. Recurrent joint pain, decrease dexterity, loss of mobility like symptoms are suffered in the later stages as the complications [4, 5, 55, 56,]. Other

complications like uveitis and retinitis [22, 30, 32] and serious ones like myocarditis, meningoencephalitis and hemorrhage are also observed [31, 36, 37, 39]. Complications are more in case of elders above 65 years of age and mortality rate is high up to 33% in this case.

The infection is confirmed with the help of different detection assay kits available commercially based on the identification of virus, viral RNA, or antibodies present in the infected blood samples. RT-PCR based detection is also done which however can be easily performed during the first week of the infection which is the acute stage of the infection having high virus titer [25]. PRNT (Plaque reduction neutralization tests) which are specific for alphaviruses are important are reliable for virus detection only problem lies with the requirement of live virus for the test [10]. However, ELISA [21] and immunofluorescent [28] based test are available which detect both anti- chikungunya IgM and IgG in the blood sample from the acute as well as convalescent stages of the infection.

With the prevalence of chikungunya in the world from the last six decades, the infection is still devoid of any effective treatment. Although trials are going on for live –immunised vaccine but it has bore no fruit till now [11]. All the treatments for chikungunya infections are symptomatic and usually bed rest, fluids and medications for relieving fever like ibuprofen and paracetamol are given. Since CHIKV is an important pathogen and is out in nature without any therapy it becomes necessary to gather and put that information into some use which might help in developing or designing any effective therapeutic against CHIKV.

Our search for an effective and essential target from the CHIKV concluded at nsP2 protease, termed to be a papain-like cysteine protease. Cysteine proteases are named on the basis of the catalytic cysteine residue which interacts with the substrate scissile carbonyl group. Among cysteine proteases majority of them form part of family which includes papain-like proteases. Papain -like protease are most versatile members in terms of their activity as well as their origin sources. They could be found in plants, animals, parasites, vertebrates, invertebrates and also in viruses [45]. nsp2CHIKV protease here belongs to C9 family of clan CA which includes non structural proteases from togaviridae [2, 44, 58].

The protease cleaves the nsP1234 polyprotein chain at three different conserved positions developing individual functional enzymes. Cleavage at these positions might need the assistance

of other non structural proteins since the *in vitro* activity studies have shown a difference in the rate of reaction [26, 41, 49, 62]. Studies with peptide substrates in case of SFV has shown that the cleavage was most efficient for 3/4 site followed by 1/2 site and negligible activity was observed in case of 2/3 peptide further affirming the cis and trans mode of nsP2 protease activity [61].

Based on this information, we were thus interested in exploiting the structural information of nsP2, an essential protease of the virus and in the development of environment friendly nonradioactive screening assay for the enzyme. Therefore, we have cloned, expressed, purified and crystallized non-structural protease nsP2 from the CHIKV and have also tried to develop a colorimetric assay by targeting this protease.

2.2 Materials (for nsP2CHIKV (471-791 amino acid)) cloning, expression and purification

2.2.1 Chemicals

PCR purification, Gel extraction and Plasmid Miniprep kits were purchased from Qiagen. Agarose, Ethidium bromide, Gel loading dye were purchased from Biochem. Agarose gel electrophoresis unit and gel imager used were from Biorad. Isopropyl β-D-1-thiogalactopyranoside used for expression purpose was purchased from Biochem.

Reagents for sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and Bradford assay were purchased from Bio-rad. Chromatography columns HiTrap SP HP (High Performance), HisTrap HP and Superdex 75 16/60 columns were purchased from GE healthcare. Centricon Ultra filtration amicons with 10 kilodalton (kDa) cut-off were purchased from Millipore. All other general chemicals were purchased either from Merck India or from Himedia chemicals.

2.2.2 Enzymes

Restriction enzymes *Nco*I, *Nhe*I, *Xho*I and for ligation T4 DNA ligase was purchased from NEB. Bovine serum albumin (BSA) and reaction buffers were from NEB. Enzymes utilized during purification, lysozymes and Deoxyrobonuclase I (DNase I) were from Sigma-Aldrich.

2.2.3 Vectors and Bacterial strains

For expression purpose *Rossetta* (DE3) cells were used since they have plasmid for the rare codons needed for viral proteins.

For developing nsP2CHIKV protease construct previously purchased pET-28c plasmid (from Novagen) with thrombin cleavage site and one pET-28c with thrombin cleavage site modified to TEV cleavage site was used. The cDNA consisting of non-structural proteins (nsP1, nsP2 and nsP3) of chikungunya was a gift from Prof. Richard J Kuhn, Purdue University.

2.2.4 Oligonucleotides

For PCR amplification, DNA oligonucleotides were designed with the help of Oligoanalyzer tool of IDT technologies. Keeping in mind that *Nhe*I and *Xho*I sites are absent in the desired

gene sequence, these sites were introduced respectively in the forward and reverse primers. The oligonucleotides were purchased from Biolinkk, Bangalore, India. For mutation, the code for the desired amino acid to be mutated was added into the complementary set of the primers designed using the above tools. The sequences of primers presented in 5' to 3' orientation (Table 2.2).

2.2.5 Culture Media

Luria bertani (LB) broth and LB agar for growing the E.coli cells were procured from Merck Speciality chemicals as dried granules. The LB culture media consisting of tryptone 1% (w/v), yeast extract 0.5%, NaCl 1% (w/v) was sterilized by autoclaving at 121 ℃ for 15 min. at 15 psi pressure. Antibiotics used for selective growth were purchased from Himedia chemicals, India.

2.2.6 Crystallization Solutions

For crystallization purpose commercially available screens i.e Crystal Screen I & II, PEG ion I & II, Index, Salt and Crystal Screen Cryo were purchased from Hampton Research. For manual optimization of conditions, highest purity chemicals were purchased from Sigma Aldrich. The solution were made in high grade nuclease free water and filtered through 0.22 µ filter (from Millipore India). The reagents were maintained at 4 ℃.

2.3 Methods

2.3.1 Cloning (without histidine tag)

pET-28c plasmid with TEV site was isolated from fresh 10mL culture of the DH5α cells containing the plasmid using Qiagen miniprep kit. PCR amplification of the desired region was done by setting up the reaction mixture and performing the reaction in the Bio-rad thermo cycler. PCR product was then purified using Qiagen pcr purification kit so as to free the pcr product from any unused dNTP's or primer-dimers. The isolated plasmid and PCR product were digested with *Nco*I and *Xho*I enzymes and incubated at 37 ℃ for 60 min. The digested plasmid and PCR products were then loaded and gel extracted from low melting agarose using Qiagen gel extraction kit. Then ligation reaction was set using the T4 DNA ligase at 16 ℃ for 15 hrs. Then ligated product was transformed into XL-1 blue cells which were spread on the LB agar

plates containing 50 µg/mL kanamycin and incubated at 37 °C for overnight. Next day colonies were screened for the clone using the PCR amplification and restriction enzyme digestion. Screened plasmids were sent for sequencing and after confirmation it was used further for nsP2CHIKV protein expression.

2.3.2 Cloning (with N terminal 6x-histidine tag)

pET-28c plasmid with thrombin site was isolated from fresh 10mL culture of the DH5α cells containing the plasmid using Qiagen miniprep kit. PCR amplification of the desired region was done by setting up the reaction mixture and performing the reaction in the Bio-rad thermo cycler. PCR product was then purified using Qiagen pcr purification kit so as to free the PCR product from any unused dNTP's or primer-dimers. The isolated plasmid and pcr product were digested with *Nhe*I and *Xho*I enzymes and incubated at 37 ℃ for 60 min. The digested plasmid and pcr products were then loaded and gel extracted from low melting agarose using Qiagen gel extraction kit. Then ligation reaction was set using the T4 DNA ligase at 16 ℃ for 15 hrs. Then ligated product was transformed into XL-1 blue cells which were spread on the LB agar plates containing 50 µg/mL kanamycin and incubated at 37 ℃ for overnight. Next day colonies were screened for the clone using the PCR amplification and restriction enzyme digestion. Screened plasmids were sent for sequencing and after confirmation it was used further for nsP2CHIKV protein expression.

2.3.3 Expression check of nsP2CHIKV (with and without 6xhis tag)

Expression plasmids were transformed into competent *E. coli* cells *Rossetta* (DE3) (made competent by $CaCl_2$ chemical method)and plated on the LB agar plates containing 50 µg/mL of kanamycin, 35 µg/mL of chloroamphenicol and incubated at 37 ℃ for overnight. 5 mL of starter cultures of cells were grown overnight at 37 ℃ at 200 rpm. Each culture was grown in a LB broth medium containing 50 µg/mL kanamycin and 35 µg/mL chloroamphenicol. The culture was used to inoculate three 100 mL of fresh LB broth-kanamycin-chloroamphenicol in a 250 mL flask. Cells were grown till O.D 600=0.4, then keeping one flask at 37 °C, two of the flasks was shifted to 25 °C and 18 °C till their O.D reached 0.6. At 0.6 O.D, all the cells were induced with 0.4 mM IPTG and incubated at 37 ℃ for 4 hrs, at 25 °C for 6 hrs and for cells

incubated at 18 °C the incubation period was for 18 hrs. After the incubation period was over cells were pellet down at speed of 3214 g for 10 min. and pellets resuspended in 5 mL of 50 mM Tris buffer (pH 7.5) were lysed using sonicator.

2.3.4 Expression Analysis

After growth, 5mL of cell culture was pelleted in a microcentifuge (3214 g for 10 min). The pellet was then resuspended in 50mM Tris buffer (pH 7.5) and DNaseI was added to final concentration of 0.01 mg/mL with 10 mM final concentration of MgCl2. The cells were then disrupted using sonicator by giving 10 pulses of 20 kHz, each of 30 seconds. The disrupted cells were then centrifuged at high speed of 28,928 g for 15 min. The soluble part referred as supernatant and the insoluble one referred as pellet, gets separated. The resulting pellet is again resuspended in 50 mM Tris buffer (pH 7.5) and all samples are then run on the SDS-PAGE along with the un-induced supernatant and pellets and protein ladder.

2.3.5 SDS-PAGE analysis

Samples were prepared for electrophoresis by mixing an appropriate amount of protein solution, usually containing 2-20 µg of protein, with 6x SDS-PAGE load buffer (0.125 M Tris pH 6.8, 9% β-Mercaptoethanol (βME), 20% glycerol, 4% SDS, 0.1 % Bromophenol blue). Samples were heated at 95 °C for 10 min. and then loaded on 12 % SDS-PAGE gels. Broad range protein ladder from GeneDireX was used molecular weight ladder for each gel. Gels were run according to the Lamelli method and were stained using staining solution containing CBB R-250 dye. Gels were then destained using destaining solution after staining for 1-2 hrs.

2.3.6 Purification of pET28c-nsP2CHIKV (without 6x histidine tag)

Optimized conditions were used for expression. Cell pellet from 1L cell culture was resuspended in 30 mL of buffer A (50 mM Tris pH 7.5, 5% Glycerol, 0.5 mM DTT) containing 0.5 mg/mL of lysozyme and 0.01 mg/mL of DNaseI. The cells were incubated for 10 min. and then lysed using French press (Constant cell Disruptor Systems, UK). The disrupted cells were then subjected to

centrifugation so as to separate the soluble portion from the insoluble fraction (at 28,928g, 60 min, 4°C). In the meantime, the Hitrap SP HP column was equilibrated with the buffer A.

The cleared cell lysate was then loaded on the equilibrated column at the flow rate of 0.4mL/min. Protein was eluted with 10x column volume linear gradient from 0% to 100% buffer B (50 mM Tris pH 7.5, 5% Glycerol, 500 mM NaCl, 5 mM DTT), at the rate of 2 mL/min followed by isocratic wash with 100% Buffer B. 5 mL fractions were collected over the course of gradient. Fractions containing the nsP2CHIKV were pooled (based on the absorbance peak) and dialysed against the dialysis buffer I (50 mM Tris pH 7.5, 5% Glycerol, 0.5 mM DTT). Sample from fractions were run on 12% SDS-PAGE gel so as to understand the protein purity. The dialysed protein was then concentrated using Amicon (10 kDa cut-off) till 1 mL final volume and loaded on pre-equilibrated (50 mM Tris pH 7.5, 5% Glycerol, 0.5 mM DTT) superdex 75 16/60 column. Protein eluted at the desired column volume was then concentrated using Amicon (10 kDa cut-off), only after confirmation by loading on the SDS-PAGE. Protein was concentrated to around 10 mg/mL.

2.3.7 Protein Concentration determination

The nsP2CHIKV concentration was determined using the Bradford Assay [20] with BSA as the Standard. A standard curve consisting of 0.2, 0.4, 0.7 and 0.9 mg/mL BSA was generated with each new set of measurements. Protein content is indicated by a change in color of the bradford dye, at 595nm, upon protein binding.

2.3.8 Purification of pET28c-nsP2CHIKV (N-terminal 6xhistidine tag)

One liter cell culture was grown in a 2 liter baffled flask at 200 rpm. Cells were grown initially at 37 °C and at O.D. 600 = 0.4 cells were shifted to 18°C of temperature. This was followed by induction with 0.4mM I.P.T.G at O.D = 0.6 after which cells were grown at the same conditions for 20 hrs. After this cells were pellet down by centrifugation at (3214 g, 10 min.) and cell pellet was resuspended in 30 mL of buffer A (50 mM Tris pH 7.5, 10 mM Imidazole, 5% Glycerol, 500 mM NaCl 0.5 mM DTT) containing 0.5 mg/mL of lysozyme and 0.01 mg/mL of DNaseI with 10 mM MgCl2. The cells were incubated for 10 min. and then lysed using French press (Constant cell Disruptor Systems, UK). The disrupted cells were then subjected to centrifugation

so as to separate the soluble portion from the insoluble fraction (at 28,928 g, 60min., 4°C). In the meantime the 5 mL His trap HP column was equilibrated with the buffer A. The cleared cell lysate was then loaded on the equilibrated column at the flow rate of 0.4mL/min. Protein was eluted with 10x column volume linear gradient from 0% to 100% buffer B (50 mM Tris (pH 7.5), 250 mM Imidazole, 5% Glycerol, 250 mM NaCl, 5 mM DTT), at the rate of 2 mL/min followed by isocratic wash with 100% buffer B. 5 mL fractions were collected over the course of gradient. Fractions containing the nsP2CHIKV were pooled and dialysed along with 20 mg/mL thrombin (1:50 thrombin:protein) against the dialysis buffer I (50 mM Tris pH 7.5, 5% Glycerol, 0.5 mM DTT). Sample from fractions were run on 12% SDS-PAGE gel to assess the purity of protein sample. The dialysed protein was then reloaded on the equlibrated His trap HP column and the 6xhis tag cleaved protein was eluted in the flow-through. The column was then washed with increasing gradient from 0% to 100% imidazole in buffer B. Cleaved protein and other fractions which are collected in the gradient are then loaded on the 12% SDS-PAGE so as to confirm the cleavage. The cleaved nsP2 CHIKV protein was then concentrated using Amicon (10 kDa cut-off) till 1 mL volume (~5mg/ mL) and loaded on pre-equilibrated (50 mM Tris pH 7.5, 5% Glycerol, 0.5 mM DTT) superdex 75 16/60 column so as to separate the protein from the protease thrombin. Protein eluted at the desired column volume was then concentrated using Amicon (10 kDa cut-off), only after confirmation by loading on the SDS-PAGE. Protein was concentrated to around 8 mg/mL.

2.4 nsP2CHIKV molecular weight determination

2.4.1 Materials

The 5 enzymes comprising the standard curve in gel filtration analysis, ribonuclease A (RNase A), ovalbumin, thyroglobulin, aldolase and aprotinin, blue dextran were purchased from GE healthcare. The gel filtration column superedex 75 16/60, was also purchased from GE Healthcare. All other chemicals were from Merck India or Sigma-Aldrich.

2.4.2 Size exclusion chromatography

In order to determine the oligomeric state of the nsP2CHIKV in the solution, an analytical

superedex 75 16/60 column was run at 4 ℃ at 0.5 mL/min, with running buffer (50 mM Tris pH 7.5, 100 mM NaCl, 5% Glycerol, 0.5 mM DTT). The void volume was first determined by injecting 50 µL of 8mg/mL of blue dextran and establishing its elution volume (40 mL). A series of standards (1mg each) were dissolved in 1000 µL of the running buffer, of which 500 µL was loaded on the column. The retention volume of each standard was recorded and their KAV values were calculated using the equation.

$$KAV = \frac{VR - VO}{VC - VO}$$

Where VR is the retention volume, V0 is the void volume and Vc denotes the geometric bed volume. A plot of KAV as a function of log molecular weight produces a standard curve. Purified pET-28c-nsP2CHIKV was diluted to 4 mg/mL in the buffer A, and 500 µL was injected on the column and run under the conditions as described above. Retention times of the peaks were recorded, and the MW's (molecular weights) of the each standard were determined using the above equation and by fitting the standard curve.

2.4.3 Native PAGE

The oligomeric state of the protein was further confirmed by running the protein on the PAGE gel under the non reducing conditions. 12% PAGE gel was set but without the use of the SDS detergent. The 5x loading dye for native PAGE gel was also lacking any reducing agents like β-mercaptoethanol or detergents like SDS.

The purified protein was loaded on the native gel along with standard protein marker and run with the voltage of 90 V. After completion of the run, the gel was stained and destained using the standard protocol.

2.5 Crystallization and data collection

Crystallization of CHIKV protease was tried by sitting drop vapor diffusion method. The N-terminal 6xhis tag of the protein was removed by cleavage with thrombin and the purified protein was concentrated to 8 mg/ mL in 25 mM Tris buffer (pH 7.5), 5% Glycerol and 100 mM NaCl, 0.5 mM DTT. Initial crystal hits were obtained in the condition containing 1.6 M sodium

citrate tribasic dihydrate (pH 6.5) at 4°C. The crystals obtained were small and not of the size to be mounted for diffraction. So to improve the quality and size of the crystals manual trials were done. 1 µL of the purified protein solution was mixed with 1 µL of well solution containing different molarities and pH of sodium citrate. The drop was equilibrated at 4°C. Mountable crystals of the nsP2CHIKV appeared after 4-5 days. Prior to mounting, crystals were harvested by a 30 sec soak in cryo protective solution containing 5% glycerol, 0.6 M sodium citrate tribasic dihydrate pH 6.0, followed by rapid submersion in liquid nitrogen. The diffraction data were collected at 100 K from a single large crystal using a a Bruker Microstar copper rotating anode X-ray generator (CuKα wavelength = 1.54 Å). The images were collected on MAR345 image plate detector. The crystal to detector distance was kept 200 mm and images were collected with exposure time of 10 min. and an oscillation width of 1° per image. The crystal was diffracted maximum up to 2.60 Å resolutions. The data were processed and scaled using HKL2000 suit [38].

2.5.1 Structure refinement and statistics

For the structure solution, the reflections were indexed, integrated, and scaled using the HKL2000 program suite [38]. The nsP2CHIKV crystal belongs to space group *P212121*, the corresponding refinement statistics are shown in Table 2.3. The initial phases for nsP2CHIKV were obtained by molecular replacement with MOLREP of the CCP4 suite [8, 60]. The protein coordinates of single subunit of nsP2CHIKV (PBD ID: 3TRK) that shares highest sequence and structure identity with our protein construct detected by the *DALI* server was used as suitable search template for molecular replacement [60]. All ligands and waters coordinates were removed from the search template. This model provided sufficient phase estimates for subsequent model building and yielded a solution with one molecule per asymmetric unit. The reflections within the resolution range 74.59-2.59 Å were selected for refinement. The rigid body refinement was followed by iterative cycles of restrained atomic parameter refinement including TLS refinement with REFMAC5 [34] and PHENIX [1]. The repetitive cycles of model rebuilding based on σA-weighted 2Fo–Fc and Fo–Fc maps were performed by employing COOT [12]. The water molecules were added in the peaks contoured at 3σ in the Fo-Fc

difference Fourier map which simultaneously satisfying density contoured at 1σ in the 2Fo-2Fc map.

2.6 Active site cysteine mutant

Active site cysteine was converted to alanine to use it as a negative control in biochemical assays and to use purified protein for the formation of protease-peptide complex for crystallization. So complementary mutant primers were designed [*sequence* given in table 2.2], and the C478A mutation was performed by site directed mutagenesis and confirmed by sequencing. The developed mutant protein was expressed and purified for various studies.

2.6.1 Material

The oligonucleotides for the mutant were designed using oligoanalyser tool of IDT technologies and were purchased from the Biolinkk. dNTP's, Phusion polymerase, *Dpn*I were purchased from the New England Biolabs. Sequencing was done by the Biolinkk.

2.6.2 Methods

Using pET-28c-nsP2CHIKV (with 6x-histidine tag) as the template, PCR reaction was set. The amplification was observed at around 7 kb size of the DNA ladder. The amplified product was subjected to the *Dpn*I digestion at 37 °C for 60 min. along with unamplified pET-28c-nsP2CHIKV as the control. After incubation, the digested products were run on the 0.8% medium melting agarose.

The amplified PCR product was observed as a single band at the 7 kb size whereas in control plasmid smear and different size fragments were observed. The 5 µL of DpnI digested product was then transformed in the XL-1 blue cells which have the property to religate the nick DNA fragments. The transformed cells were spread on the LB agar plates containing 50 µg/mL kanamycin and incubated at 37 °C for overnight. Next day, the observed colonies were screened for the mutant amplified pET-28c-nsP2CHIKV plasmid, which was confirmed by restriction digestion with *Nhe*I and *Xho*I restriction enzymes. The confirmed plasmid was then sent for sequencing to confirm the cysteine mutation.

2.6.3 Expression and purification of pET-28c-nsP2CHIKV Cys478Ala mutant

The nsP2 CHIKV Cys478Ala mutant was expressed and purified with the affinity chromatography and size exclusion method as described in the pET-28c-nsP2CHIKV (with 6xhis tag).

2.7 TNBS activity assay and kinetic studies

2.7.1 Material

The amino acid sequences of CHIKV nsP2 substrate sites RAGAGIIE [P1], RAGCAPSY [P2], RAGGYIFS [P3] were designed and purchased from GL biochem systems. The peptides were purified by HPLC and confirmed by mass spectrometry (data provided by the company0. The peptides were water soluble and thus can be easily used for the assay. Phosphate buffer, 2,4,6-trinitrobenzene sulfonic acid (TNBS) were purchased from Sigma (USA).

Activity Buffer: Sodium Phosphate (pH 6 to 9) and 50 mM concentration NaCl. pH of the buffer was adjusted from pH 6 to 9 using HCl and NaOH. The substrate peptide stocks of 1mM were made in the activity buffer only.

2.7.2 Methods

The assay was performed using the purified protein at the concentration of 200 uM. Synthetic substrate peptide stocks were made in the 50mM sodium phosphate buffer, pH 7.5. The purified protein was incubated with substrate in the presence of 50mM sodium phosphate buffer, 150 mM NaCl and the reaction mixture was incubated at 30 °C in water bath for 60 min. 2 µL of 0.1% TNBS was then added into the mixture and reaction was incubated at 50 °C in dark environment for color to develop. Reaction was then stopped with 5 mL of 0.1N HCL addition. The absorbance was then taken at 420nm along with blanks [14, 16].

2.7.3 Kinetic Analysis of nsP2CHIKV Protease and its Mutant C478A

The kinetic studies were done by varying the reaction time in the TNBS reaction described above. Kinetic constants *k*cat and *K*m were determined for the wild type and the mutant protease. Effect of pH on the activity was also studied in case of wild and mutant protease by

using buffers of pH range 5 to 9. Change was observed in case of wild type in contrast to the mutant protease for which no major change in absorbance has been observed at all of the pH range. In case of reaction in the presence of substrate peptides the mutant protease did not show any considerable absorbance whereas major shift in case of wild type protease was observed.

2.8 Results and Discussion

2.8.1 pET-28c-nsP2CHIKV (6x histidine tagged and untagged) cloning, expression and purification

The wild pET-28c-nsP2CHIKV was cloned in the pET-28c vector between *Nco*I and *Xho*I sites, without any 6xhis tag to skip the tag cleavage step. The viral protein sequence of 475-798 of nsP2 protein, containing the protease domain, was cloned in the plasmid pET-28c. Protein expressed in pET series of vectors is expressed under the control of T7 promoter; the T7 polymerase is provided by the host expression cells which are λDE3 lysogens. Cloned in pET-28c between *Nco*I and *Xho*I sites, the protein is expressed as a native, non-taggged protein, whereas pET-28c [with thrombin site] constructs cloned between *Nhe*I and *Xho*I expresses N terminal 6xhis-tagged protein.

Table 2.2. Oligonucleotides used in cloning of nsP2CHIKV and for making C478A mutant.

Gene name	Insertion Vector	Primer direction	Primer sequence	Incorporated restriction enzyme site
nsP2CHIKV (without 6xHis tag)	pET28c	F	GATTCTCCATGGATACATTCCAAA ATA AAGCCAACG	*NcoI*
		R	GATTCTCTCGAGTTAGAAGGCTGCATTCAGTTGATTG	*XhoI*
nsP2CHIKV (with 6xHis tag)	pET28c	F	GATTCTCATATGATACATTCCAAA ATA AAGCCAACG	*NheI*
		R	GATTCTCTCGAGTTAGAAGGCTGCATTCAGTTGATTG	*XhoI*
nsP2CHIKV (with 6xHis tag) C478A mutant	pET28c	F	TAAAGCCAACGTTGCGTGGGCTAAGAGCTTGG	-
		R	CCAAGCTCTTAGCCCACGCAACGTTGGCTTTA	-

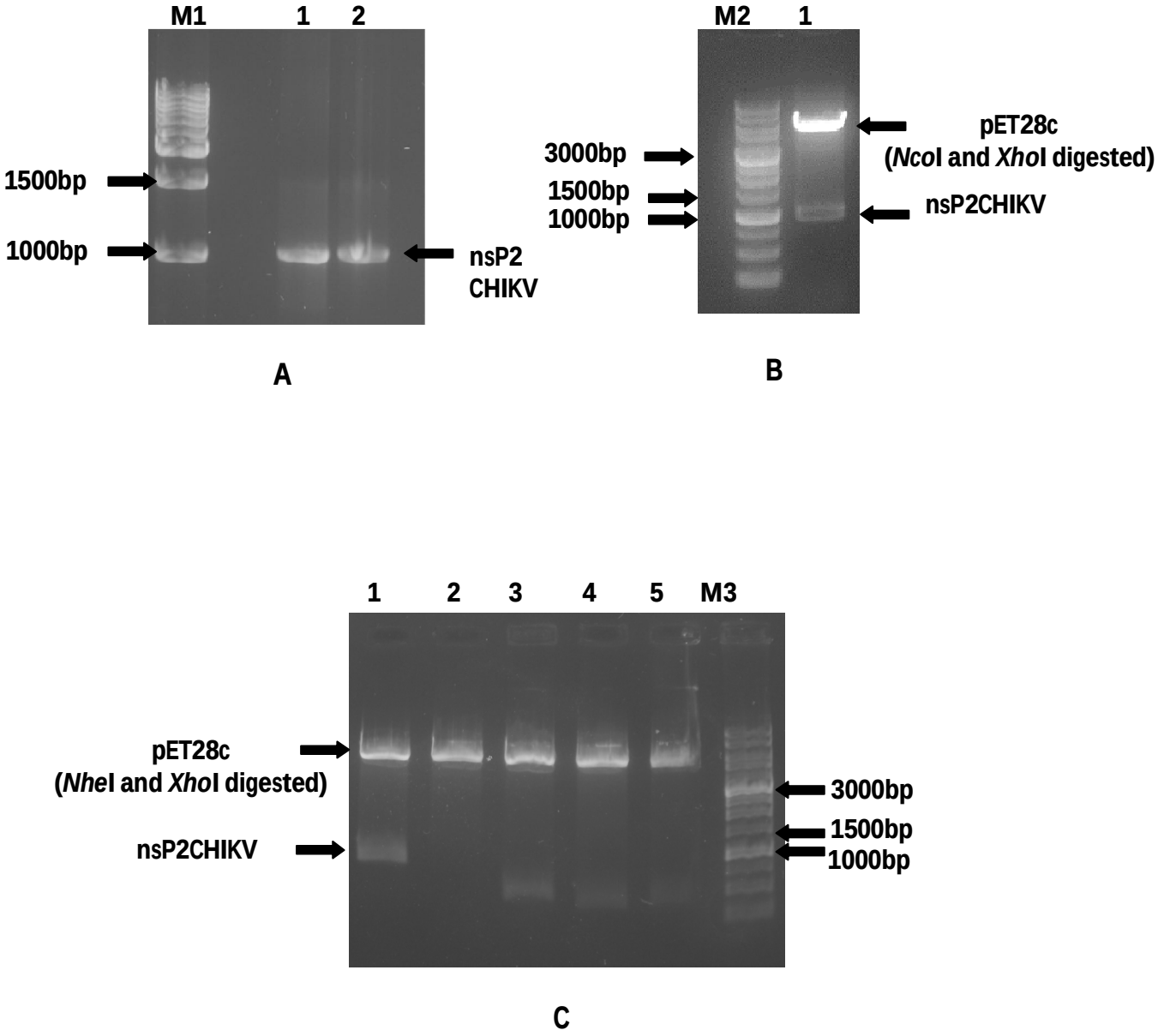
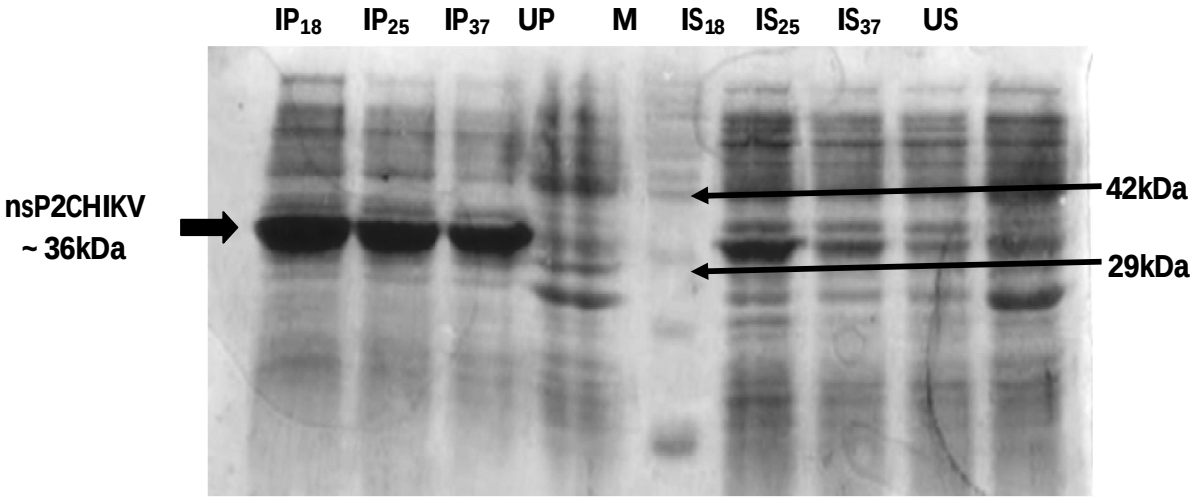


Figure 2.3 Amplification and cloning of nsP2CHIKV (with 6xhis tag and without 6xhis tag) gene. Agarose gel electrophoresis **(A)** PCR products. lane M1: DNA marker; lane 1: PCR product amplified from CHIKV genomic cDNA with nsP2CHIKV (without 6xhis tag) gene specific primers, lane 2: PCR product amplified from CHIKV genomic cDNA with nsP2CHIKV [with 6xhis tag] gene specific primers. **(B)** Recombinant plasmid digested with *Nco*I and *Xho*I. lane M2: DNA marker; lanes 1: pET-28c-nsP2CHIKV (without 6xHis tag) from screened colony. **(C)** Recombinant plasmids digested with *Nhe*I and *Xho*I. lane M3: DNA marker; lanes 1-5: pET-28c-nsP2CHIKV (with 6xhis tag) from screened colony.

Expression of both of the constructs, which were transformed into *Rossetta* (DE3) cells, was induced with IPTG at 18 °C for 20 hrs. Although expression was there at higher temperatures but soluble protein was observed at lower temperature conditions i.e 18 °C (Figure 2.4 A).

Purification of non-tagged pET-28c-nsP2CHIKV protein was pursued so as to avoid potential step of cleaving the 6xhis tag. Typically, 2 litres of cells were used per purification. Cells were lysed after gentle treatment with lysozyme followed by French press. Protease inhibitor cocktails were not used so as to avoid any effect on the protease activity of the enzyme, and hence to maintain the stability of the protein all the process were performed at 4°C.

After disruption high speed centrifuge, with fixed angle rotors were used so as to separate the debris from the soluble fraction containing the protein. After this, the protein was loaded on a 5 mL Hi trap SP column and eluted with a NaCl gradient. The calculated pI for the non-tagged pET-28c-nsP2CHIKV was found to be 9.1, indicating it would be positively charged at pH 7.5 and thus should bind to a cation exchange column. Protein obtained after this step was observed to be pure but to get >90% pure protein next step size exclusion chromatography was done.



A

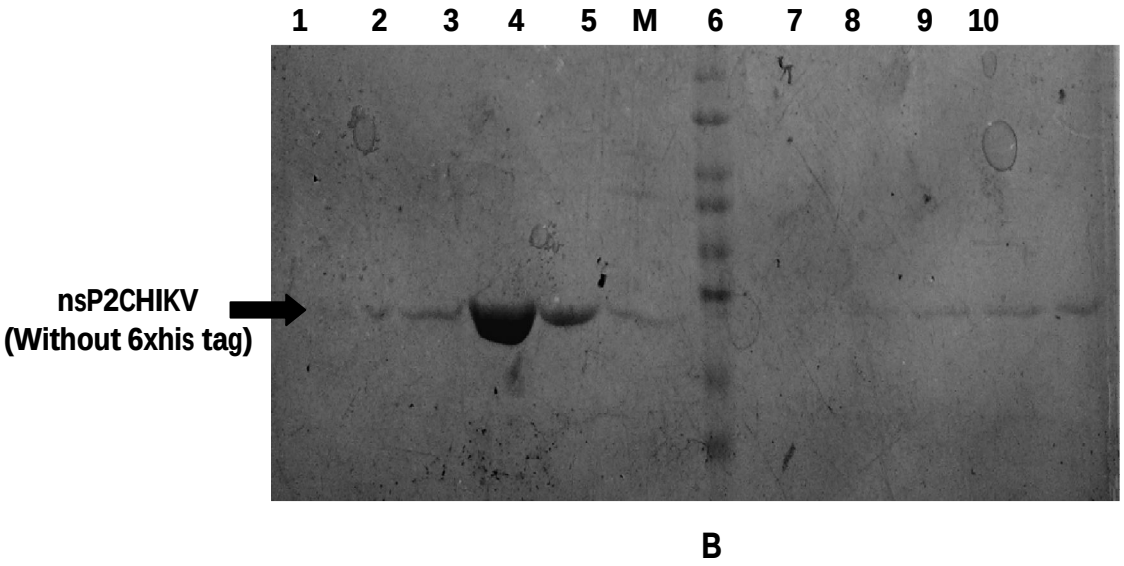


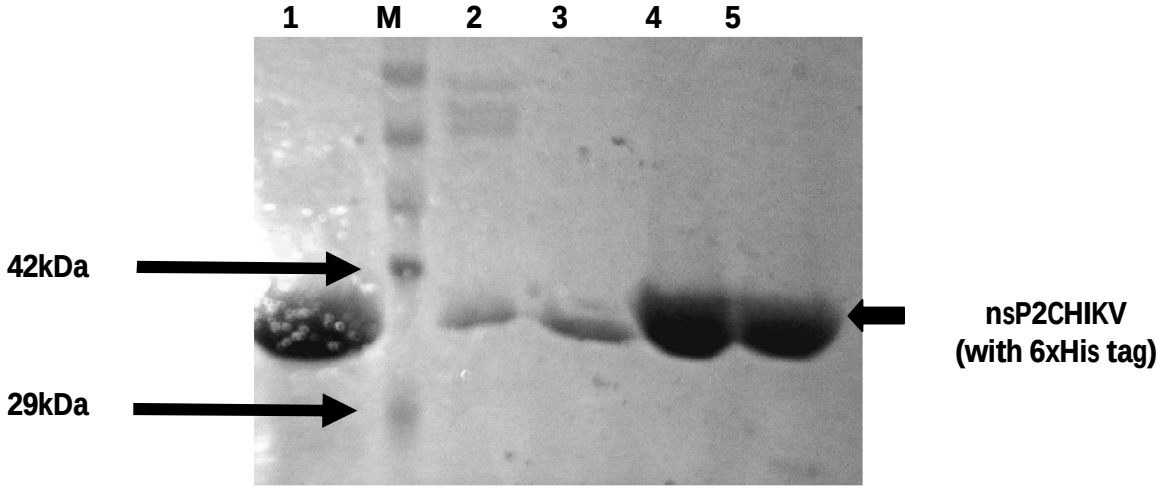
Figure 2.4 Optimized expression and purification of recombinant nsP2CHIKV (without 6xhis tag) (A) Expression analysis at different temperature conditions and soluble expressed recombinant protein was obtained at 18 °C. **(B)** Purification of nsP2CHIKV (without 6xHis tag) using Hitrap SP column. Lane M: pre-stained protein ladder; Lane 1-10: purified fractions.

Since till now monomeric states of the protein in case of the other alphaviruses is known, so use of superdex 75 16/60 sephadex column was considered appropriate. The monomeric protein eluted from the column was observed on the 12% SDS-PAGE gel and observed pure. Purification from 2 litre culture pellet yields around 100 µL of protein with approximately 6 mg concentrations (calculated by Bradford assay).

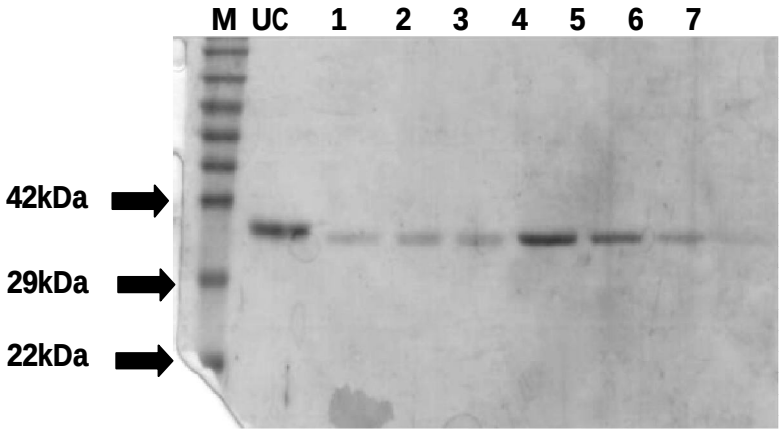
The protein was used for crystallization trails by using the Hampton crystal screen. Protein was tried to crystallize using hanging drop vapor diffusion method. The crystal tray was set using Hampton crystal screen 1-2, peg ion 1-2, natrix 1-2, salt screen 1-2, and crystal screen cryo 1-2. Protein was set for crystallization with the reservoir at 1:1 protein: reservoir at 4°C, and 20°C.

But since no precipitation was observed for long time in the drops it was assumed that we need to increase the concentration for obtaining the optimum amount needed for crystallization.

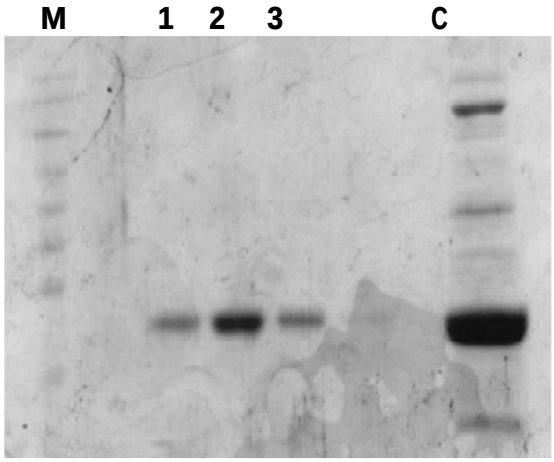
Although we were able to purify nsP2CHIKV without tag also, but some time not being able to get pure and enough protein and low yield by this method paved way for cloning the protein with the 6xhis tag. The pET-28c-nsP2CHIKV (with tag), was grown in the same manner as for the non-tagged protein. After culture the first step of purification was affinity chromatography based on the affinity of 6xhis tag towards the Ni^{2+} present in the Histrap HP column (Figure 2.5 A).



A



B



C D

Figure 2.5 Optimized purification of recombinant nsP2CHIKV (with 6xHis tag) (A) Affinity purification of nsP2CHIKV (with 6xHis tag) using Histrap columns. lane M: prestained protein ladder; lane 1-5: purified protein fractions. **(B)** Reverse NTA step of purification of nsP2CHIKV after the cleavage of 6xHis tag using Thrombin. lane M: prestained protein ladder; lane UC: uncleaved nsP2CHIKV protein as the control; lane 1-7: nsP2CHIKV fractions after the 6xHis tag cleavage.**(C)** Purified cleaved nsP2CHIKV fractions obtained from [B] applied to a Superdex-75 gel filtration column. lane M: prestained protein ladder; lane C: concentrated cleaved nsP2CHIKV protein; lane 1-3: fractions obtained after superdex-75 purification. **(D)** Concentrated purified nsP2CHIKV after final step of purification.

Followed by purification, the protein was dialyzed and at the same time incubated with thrombin so as to cleave the 6xhis tag from the protein. The thrombin site is present between the 6xhis tag and gene expression sequence in pET-28c. The incubated protein was again loaded on the column and the flow through consists of the protein with tag cleaved (Figure 2.5 B). The protein is concentrated up to 1 mL with 15 mg/mL concentration using Amicon (10 kDa cut-off) concentrator and then loaded on the superdex 75 16/60 column (Figure 2.5 C). Although pure protein is attained by the affinity and after cleavage step but for separating the thrombin protein from the nsP2 protease this step is necessary.

2.8.2 Oligomeric State Determination

2.8.2.1 Size exclusion chromatography

To determine the oligomeric state of nsP2CHIKV, size exclusion chromatography on superdex 75 16/60 column was performed which was standardized previously with 5 proteins of known molecular weight (Figure 2.6B). The nsP2CHIKV eluted as a single peak from the column whose molecular mass when calculated would come around 39 kDa which is close to the mass calculated on the basis of the sequence of the protein (Figure 2.6A).

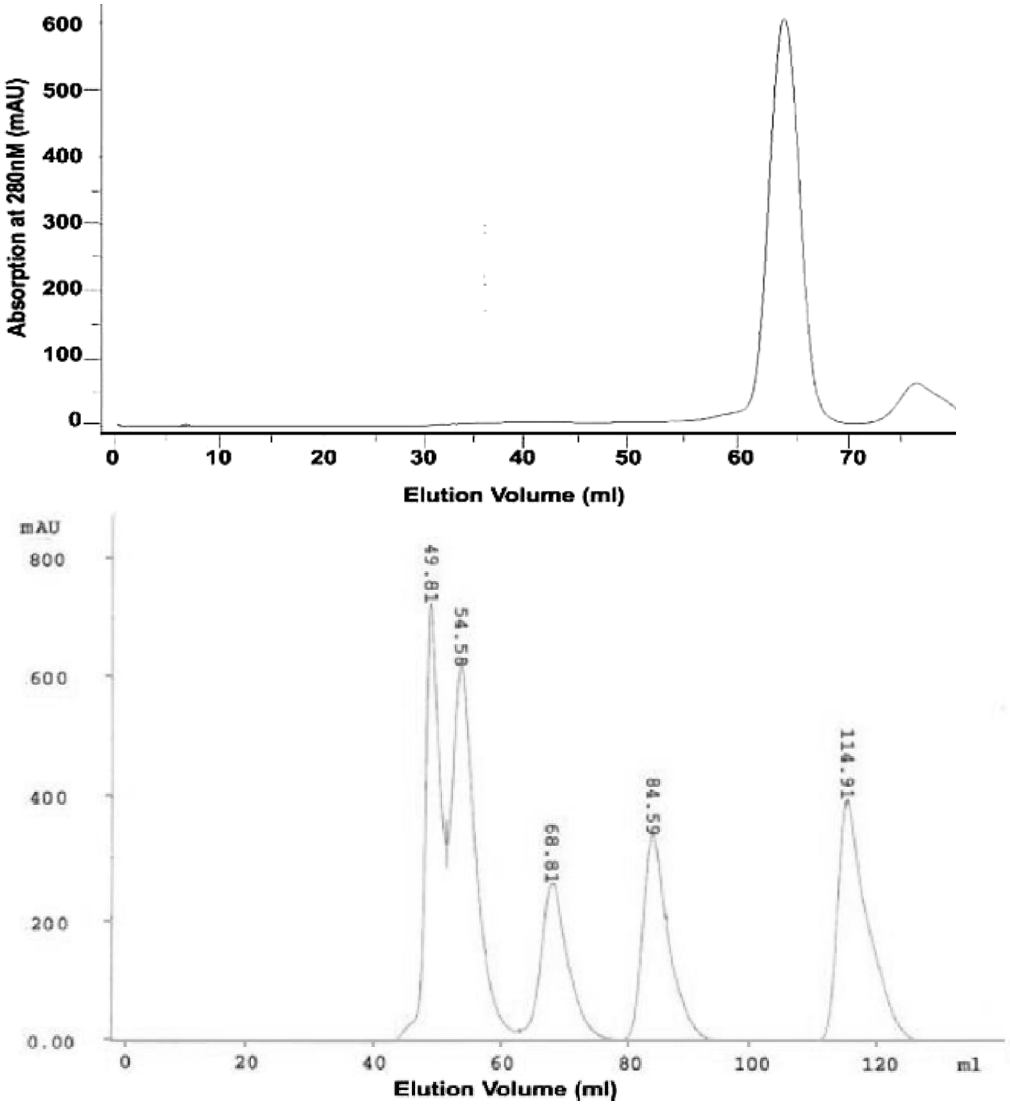


Figure 2.6(A). Chromatogram obtained after loading the nsP2CHIKV protein on Superdex-75 16/60 column. The void volume of the column is 40 mL and after injection the

protein peak came at around 65 mL **(B) Chromatogram for standard proteins run on superdex 75 16/60 column with their molecular weights and elution volume**. Vitamin B12 (1.35 kDa) (114.91 mL), RNase (13.7 kDa) (84.59 mL), ovalbumin (44 kDa)(68.81mL), aldolase (158 kDa)(54.58 mL) and thyroglobulin (670 kDa)(49.81mL).

2.8.2.2 Native PAGE

Further confirmation of the oligomeric state of the protein was done by analyzing the protein on the PAGE gel under the non reducing conditions. Pure, single band of approximately 36 kDa was observed on the gel (Figure 2.7) which further attests to our observation that the protein is present in the monomeric state in the solution during purification.

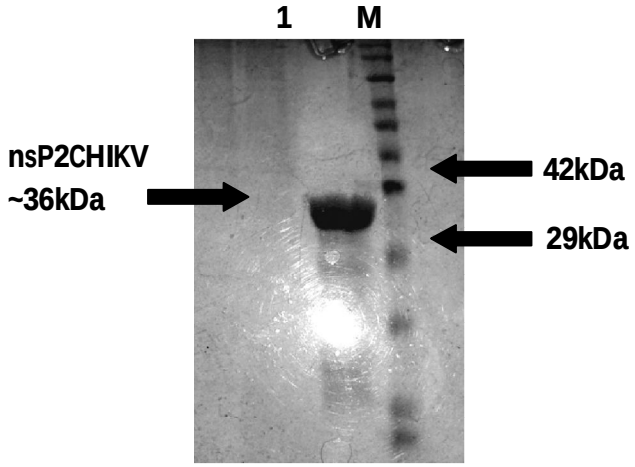


Figure 2.7 Native PAGE gel for nsP2CHIKV confirming its monomer state in solution. Lane 1- nsP2CHIKV, lane M – prestained protein ladder

2.8.3 Crystallisation and structure determination

Commercial sparse matrix screens from Hampton research, USA were utilized for the initial crystal screening. Each screen had 96 different solutions varying in precipitant, salt and additives combination. Different buffers were used to screen a wide pH range (4.0-10.0). The

concentration of nsP2CHIKV was kept at 8 mg/ mL for the initial screening. Initial hits of small crystals with three dimensional shapes were observed in a condition consisting 1.6 M sodium citrate tribasic dihydrate pH 6.5. The subsequent trial with grid screening of nsP2CHIKV around the initial condition resulted in the growth of some large diffraction quality crystals (Figure 2.8). These crystals were grown in full size over the course of 15-20 days and diffracted up to around 3 Å in initial test on X-ray home source. The data set with 99% completeness at 2.6 Å resolution was collected after screening of several crystals (Figure 2.9). The phases were calculated by molecular replacement method using MolRep utilizing a monomer of the nsP2 structure of CHIKV (PDB code: 3TRK) as the search template [60]. This model provided sufficient phase estimates for subsequent model building and yielded a solution with four molecules per asymmetric unit. The reflections within the resolution range 47.21-2.59 Å were selected for refinement. The rigid body refinement was followed by iterative cycles of restrained atomic parameter refinement including TLS refinement with REFMAC5 [34] and PHENIX [1].

The repetitive cycles of model rebuilding based on σA-weighted 2Fo–Fc and Fo–Fc maps were performed by employing COOT [12].The water molecules were added in the peaks contoured at 3σ in the Fo-Fc difference Fourier map which simultaneously satisfying density contoured at 1σ in the 2Fo-2Fc map. After the starting round of refinement without any model rebuilding, the values of Rwork and Rfree were 32.3% and 39.5% respectively. To check the error, all water molecules were inspected manually. Iterative cycles of refinement assessed by R_{free} and addition of solvent molecules yielded a model with final value of R_{work} of 18.0% and R_{free} of 24.0% (Table 2.3). The steriochemical attributes of refined model was validated by using the program PROCHECK [23] and refinement statistics are also presented in Table 2.2. The statistics of the Ramachandran plot distribution shows 92% residues in the most favored regions with 1.1% in allowed region. The average B-factor for the entire main chain residue was 55.80. The final model contained 1284 protein residues and 298 water molecules. The crystal belong to *P212121* space group with unit cell parameters a= 87.04, b = 158.96, c= 158.88, and α=β=γ=90°. The calculated Matthews coefficient assumed four molecules per asymmetric unit with a solvent content of 37%. The data collection statics are given in table 2.3.

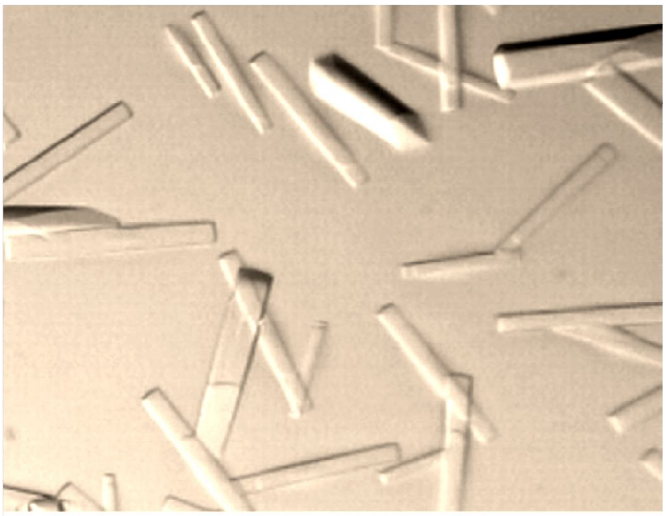


Figure 2.8 Crystals of nsP2CHIKV obtained by the sitting-drop vapor-diffusion method using 0.9 M Sodium citrate tribasic dihydrate pH 6.0 as precipitant at 4°C.

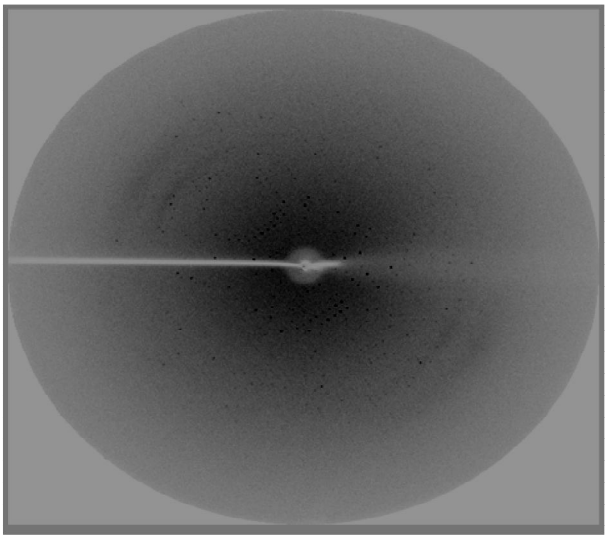


Figure 2.9 X-ray diffraction from a single crystal of nsP2CHIKV.

Table 2.3 Data processing and refinement statistics

Wavelength (Å)	1.54
Resolution range (Å)	47.21 - 2.595 (2.687 - 2.595)
Space group	*P212121*
Unit cell dimensions *α, β, γ* *a, b , c*	 87.04, 158.96, 158.88 90, 90, 90
Total reflections	165501 (2376)
Unique reflections	56663 (2141)
Multiplicity	2.9 (1.2)
Completeness (%)	82.22 (31.65)
Mean I/sigma(I)	4.43 (0.52)
R-merge	0.2257 (1.023)
R-work	0.1851
R-free	0.2503
Protein residues	1284
Water	298
RMSD from ideal geometry *Bond length* *Bond angles*	 0.010 1.35
Ramachandran plot *Ramachandran favored (%)* *Ramachandran outliers (%)*	 92 1.1
Average B-factor	55.80

Statistics for the highest-resolution shell are shown in parentheses.

2.8.4 Cloning and Purification Analysis of nsP2CHIKV active site Mutant Cys478Ala

For using as a negative control active site cysteine of nsP2CHIKV enzyme was substituted with alanine (Figure 2.10). The mutation was confirmed by sequencing and the mutant protein was expressed and purified by the protocol used for the wild type nsP2CHIKV. The protein was soluble at 18 °C after with 0.4 mM IPTG and the method for purification was same as that of the wild protein (Figure 2.11). The mutant protein was used in the activity studies and activity was observed to be abolished due to the mutation (section 2.8).

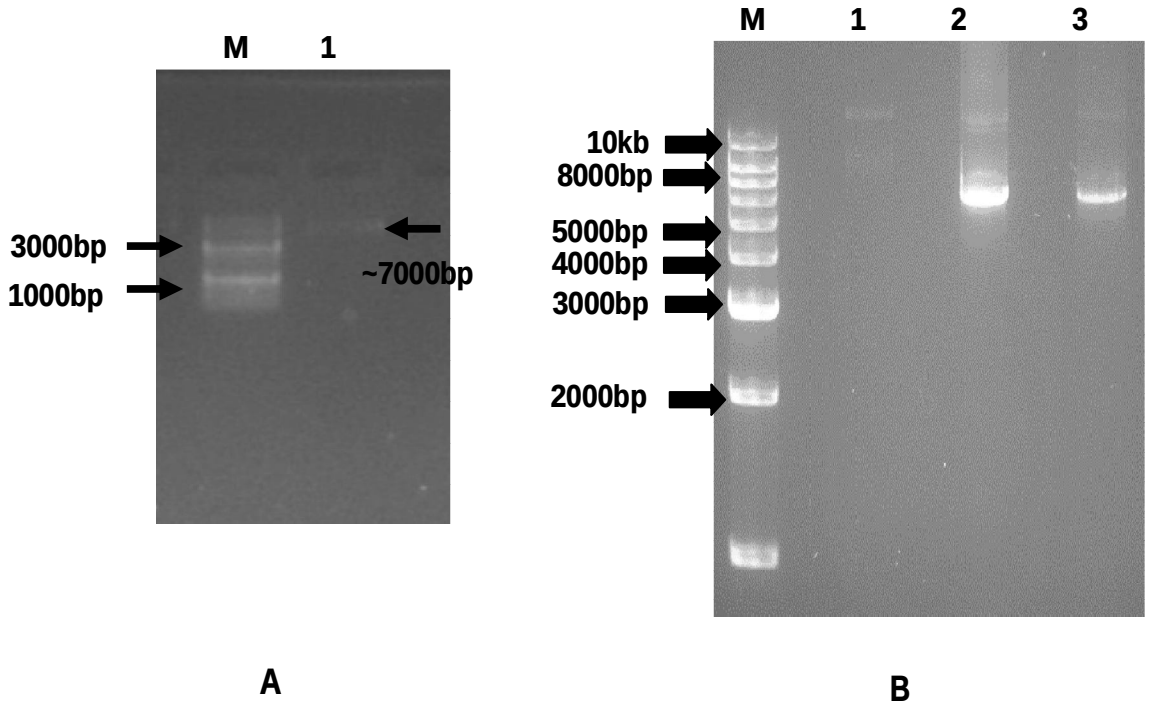


Figure 2.10 Development of C478A mutant of nsP2CHIKV. (A) PCR amplification of the pET-28c-nsP2CHIKV plasmid using the mutant primers. lane M: 1kb DNA ladder;lane 2: PCR amplified pET-28c-nsP2CHIKV. **(B)** *Dpn*I digestion of the amplified pET-28c-nsP2CHIKV-C478A plasmid. lane M: 1kb DNA ladder; lane 2, 3: *Dpn*I digested pET-28c-nsP2CHIKV-C478A.

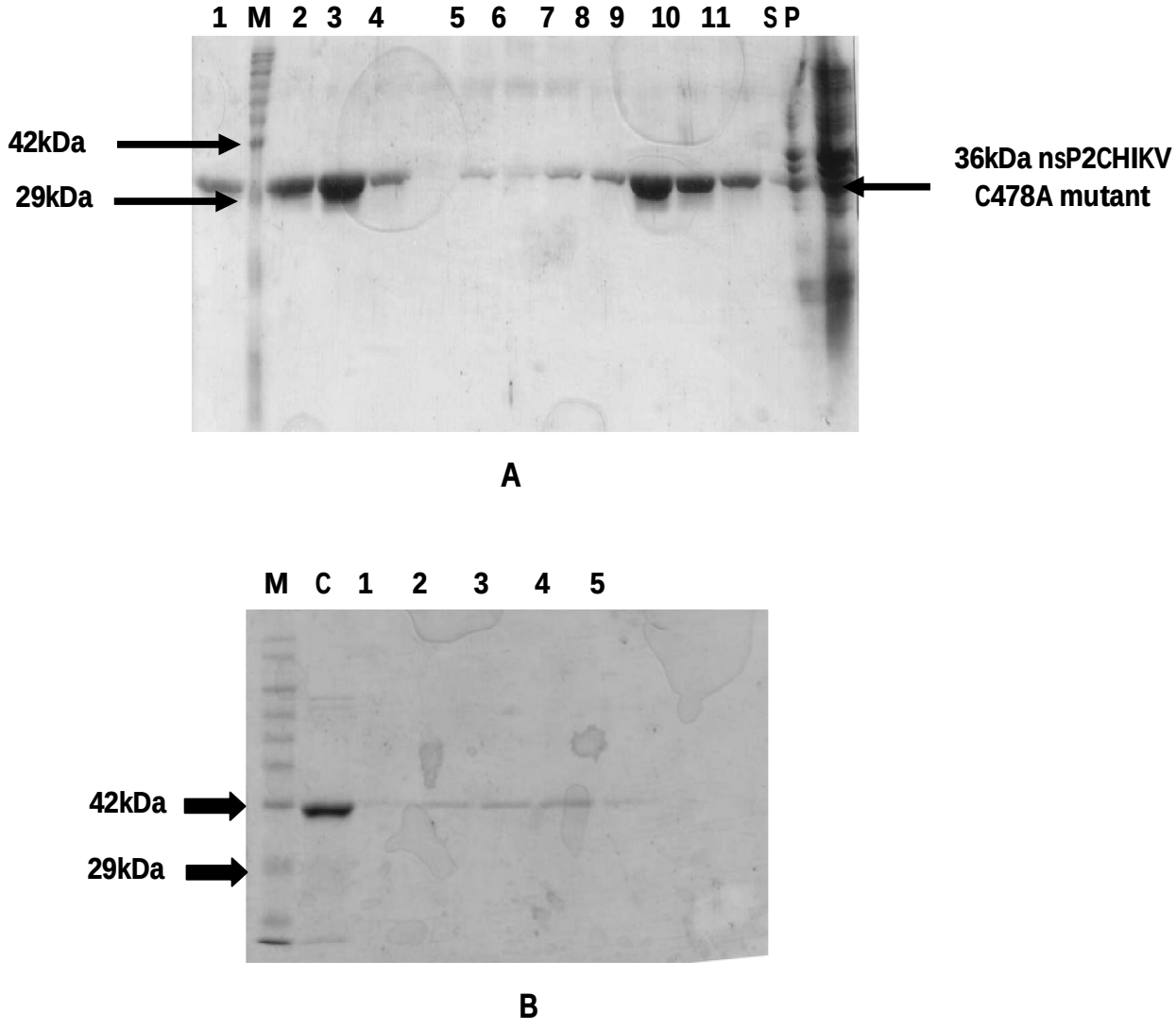


Figure 2.11 Purification of nsP2CHIKV C478A mutant. (**A**) Affinity chromatography purification of nsP2CHIKV C478A mutant using histrap column on the basis of affinity of the 6xhis towards the nickel in the histrap column. lane M: prestained protein ladder; lane 1-10: purified protein fractions; lane S: supernatant; lane P: Pellet. (**B**) Second step of purification of mutant nsP2CHIKV using superdex-75 16/60 column. lane M: prestained protein ladder; lane C: concentrated mutant nsP2CHIKV from step (**A**); lane 1-5: purified fractions from the size exclusion chromatography column.

2.8.5 TNBS activity assay

The assay was performed after optimizing different physiological conditions to identify the ideal condition for the nsP2CHIKV protease activity. The TNBS which primarily interact with the terminal amine group to form an orange colored adduct (Figure 2.12) whose intensity is measured at the 420nm in the visible range on the spectrophotometer [14, 16].

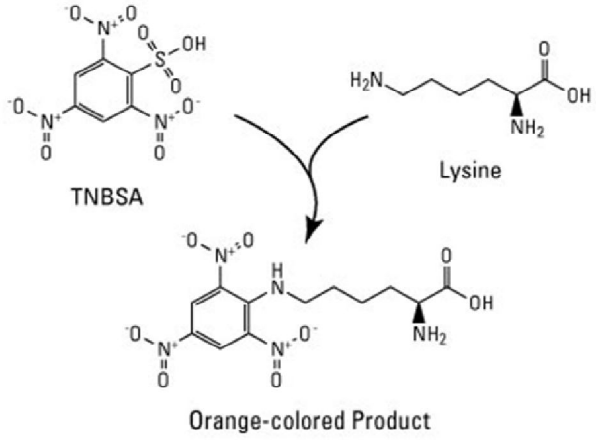


Figure 2.12 Mechanism of TNBS activity

The increase or decrease in the intensity is linked with the free amine group available in the reaction solution. When the protease acts on the substrate peptides, due to its efficiency of cleavage free amine groups are released after every cleavage, which interacts with the TNBS and helps us in the determination of the rate of the reaction with which the protease is acting on the substrate. Following this basic principle the assay was developed after optimizing the ideal nsP2CHIKV protease activity conditions (Figure 2.13).

The wild nsP2CHIKV protease was found to be behaving actively at temperatures between 25°C to temperature 37 °C. Above these temperature ranges the activity of the protein was submissive in comparison to these above temperatures. Although we did not play around with salt concentrations but we did keep 150 mM NaCl in the activity buffer, absence of which did affected the activity of the enzyme. The buffer used was phosphate for the reaction because in the presence of tris buffer the TNBS would have interacted with the amine in the tris and given us false positive readings.

The reaction was performed for both the wild and mutant nsP2CHIKV protease in the presence of substrate peptides. In comparison to the mutant C478A nsP2CHIKV, the wild type was more active. In case of wild type protease there has been increase in the absorbance to a peak with increasing pH which then showed a decreasing when moved towards alkaline range. In case of mutant protease no major change in absorbance has been observed at all of the pH range. However the rate of reaction was very high in case of peptide with nsP3/4 site but significant activity was also observed in case of nsP1/2 site peptide substrate. But no activity was observed in case of nsP2/3 peptide substrate (Figure 2.14). This could be due to the fact that for the cleavage at this site may be nsP2CHIKV needs help of nsP3 or other nonstructural proteins or could be the reason that the nsP2/3 peptide is not being able to access nsP2CHIKV protease in the conformation it is present now.

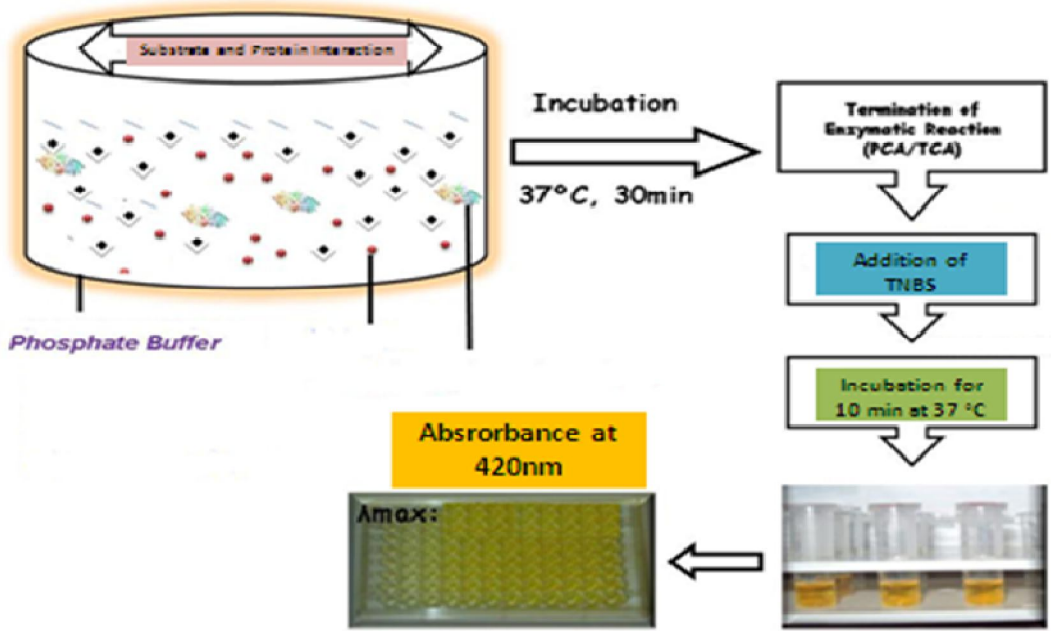


Figure 2.13 Step wise mechanism of the TNBS assay used for demonstrating nsP2CHIKV protease activity

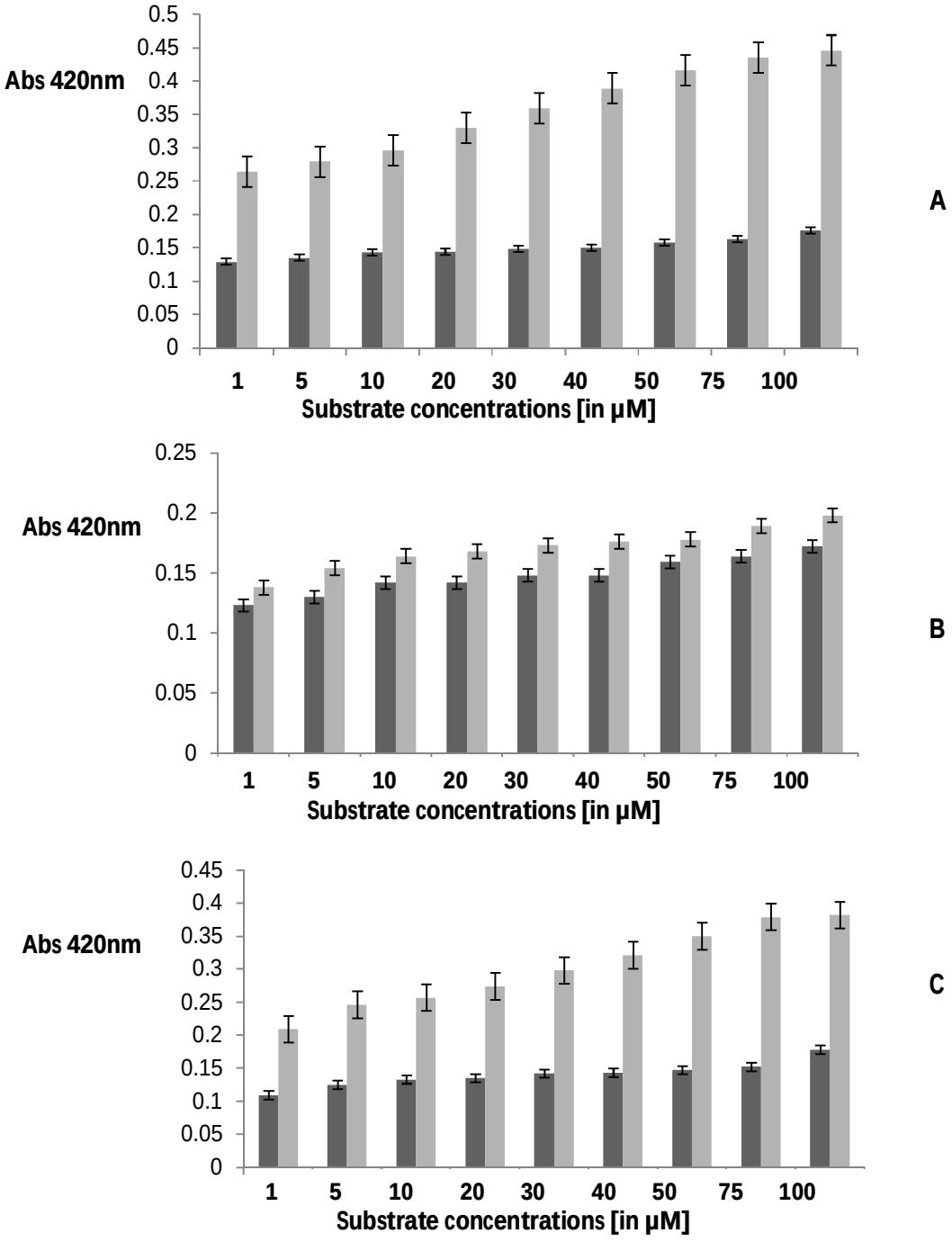


Figure 2.14 TNBS activity assay in the presence of nsP2CHIKV wild and nsP2CHIKV C478A mutant. (A) Activity in the presence of substrate peptide nsP3/4 (B) Activity in the presence of substrate peptide nsP2/3 (C) Activity in the presence of substrate peptide 1/2. Represents wild nsP2CHIKV ; represents C478A mutant nsP2CHIKV .

Substrate	**Site**	K_m**(µM)**	k_{cat} **(x** 10^6 s^{-1}**)**
RAGAGIIE	nsP1/2	1798	735
RAGGYIFS	nsP3/4	223.1	301

The kinetic studies of the nsP2CHIKV in the presence of nsP1/2 and nsP3/4 substrate peptides were also performed. The rate of reaction in case of nsP3/4 peptide was found to be high in comparison to the nsP1/2 peptide. This clearly shows the preference of nsP2CHIKV towards nsP3/4 site over the nsP1/2 site.

2.9 Conclusion

The protease domain of the nsP2 of chikungunya virus was cloned in pET-28c vector and heterologous expression of the protein was done in *Rossetta* cells. The expressed protein was purified upto 99% homogeneity by a three step purification protocol involving Affinity purification followed by the cleavage of the 6xHis tag using thrombin enzyme and in final step size exclusion chromatography was used to separate the protein from thrombin. The purified protein was put onto x=crystallization trials using the commercial Hampton screens via sitting drop vapour diffusion method. Initial hit of the tiny nsP2CHIKV crystals provided us the possible condition suitable for the crystallization. Finally we were able to develop diffracting and mountable crystals which gave us the X-ray data for nsP2CHIKV protease domain. The structure was solved using molecular replacement at a resolution of 2.6 Å. Also for screening the inhibitory molecules against the nsP2 protease, we have developed a calorimetric based assay.

2.10 Bibliography

1. Adams, Paul D., Pavel V. Afonine, Gábor Bunkóczi, Vincent B. Chen, Ian W. Davis, Nathaniel Echols, Jeffrey J. Headd et al. "PHENIX: a comprehensive Python-based system for macromolecular structure solution." Acta Crystallographica Section D: Biological Crystallography 66, no. 2 (2010): 213-221.
2. Barrett, Alan J., J. Fred Woessner, and Neil D. Rawlings, eds. Handbook of proteolytic enzymes. Vol. 1. Elsevier, 2012.
3. Borgherini, Gianandrea, Patrice Poubeau, Frederik Staikowsky, Manuella Lory, Nathalie Le Moullec, Jean Philippe Becquart, Catherine Wengling, Alain Michault, and Fabrice Paganin. "Outbreak of chikungunya on Reunion Island: early clinical and laboratory features in 157 adult patients." Clinical Infectious Diseases 44, no. 11 (2007): 1401-1407.
4. Borgherini, Gianandrea, Patrice Poubeau, Annie Jossaume, Arnaud Gouix, Liliane Cotte, Alain Michault, Claude Arvin-Berod, and Fabrice Paganin. "Persistent arthralgia associated with chikungunya virus: a study of 88 adult patients on reunion island." Clinical Infectious Diseases 47, no. 4 (2008): 469-475.
5. Brighton, S. W., O. W. Prozesky, and A. L. De La Harpe. "Chikungunya virus infection." A retrospective study of 107 (1983): 313-315.
6. Chretien, Jean-Paul, and Kenneth J. Linthicum. "Chikungunya in Europe: what's next?." The Lancet 370, no. 9602 (2007): 1805-1806.
7. Couderc, Thérèse, Fabrice Chrétien, Clémentine Schilte, Olivier Disson, Madly Brigitte, Florence Guivel-Benhassine, Yasmina Touret et al. "A mouse model for Chikungunya: young age and inefficient type-I interferon signaling are risk factors for severe disease." PLoS Pathog 4, no. 2 (2008): e29.
8. Collaborative, Computational Project. "The CCP4 suite: programs for protein crystallography." Acta crystallographica. Section D, Biological crystallography50, no. Pt 5 (1994): 760.
9. Deller, Lt Col John J., and Lt Col Philip K. Russell. "Chikungunya disease." The American journal of tropical medicine and hygiene 17, no. 1 (1968): 107-111.
10. Department of Health and Human Services; Centers for Disease Control and Prevention; National Institutes of Health. Biosafety in Microbiological and Biomedical Laboratories

(BMBL) 5th Edition. Washington DC: US Government Printing Office;2007. Arboviruses and Related Zoonotic Viruses.

11. Edelman, R., C. O. Tacket, S. S. Wasserman, S. A. Bodison, J. G. Perry, and J. A. Mangiafico. "Phase II safety and immunogenicity study of live chikungunya virus vaccine TSI-GSD-218." The American journal of tropical medicine and hygiene 62, no. 6 (2000): 681-685.
12. Emsley, Paul, and Kevin Cowtan. "Coot: model-building tools for molecular graphics." Acta Crystallographica Section D: Biological Crystallography 60, no. 12 (2004): 2126-2132.
13. Gérardin, Patrick, Georges Barau, Alain Michault, Marc Bintner, Hanitra Randrianaivo, Ghassan Choker, Yann Lenglet et al. "Multidisciplinary prospective study of mother-to-child chikungunya virus infections on the island of La Reunion." PLoS Med 5, no. 3 (2008): e60.
14. Goodwin, Jesse F., and Siu-Ying Choi. "Quantification of protein solutions with trinitrobenzenesulfonic acid." Clinical Chemistry 16, no. 1 (1970): 24-31.
15. Griffin, Diane E. in Fields Virology 5th edn (eds Knipe, D. M. & Howley, P. M.) 1023–1066 (Lippincott Williams & Wilkins, Philadelphia, 2007).
16. Habeeb, AF Sa A. "Determination of free amino groups in proteins by trinitrobenzenesulfonic acid." Analytical biochemistry 14, no. 3 (1966): 328-336.
17. Halstead, S. B., S. Udomsakdi, and P. Singharaj. "Dengue and chikungunya virus infection in man in Thailand." (1962): 1964-111.
18. Halstead, Scott B., Suchinda Udomsakdi, Pricha Singharaj, and Ananda Nisalak. "Dengue and Chikungunya Virus Infection in Man in Thailand, 1962–1964 III. Clinical, Epidemiologic, and Virologic Observations on Disease in Non-Indigenous White Persons." The American journal of tropical medicine and hygiene 18, no. 6 (1969): 984-996.
19. Higashi, N., A. Matsumoto, K. Tabata, and Y. Nagatomo. "Electron microscope study of development of Chikungunya virus in green monkey kidney stable (VERO) cells." Virology 33, no. 1 (1967): 55-69.
20. Kruger, Nicholas J. "The Bradford method for protein quantitation." In Basic protein and peptide protocols, pp. 9-15. Humana Press, 1994.
21. Lakshmi, Vemu, Mamidi Neeraja, M. V. S. Subbalaxmi, M. M. Parida, P. K. Dash, S. R. Santhosh, and P. V. L. Rao. "Clinical features and molecular diagnosis of Chikungunya fever from South India." Clinical infectious diseases46, no. 9 (2008): 1436-1442.

22. Lalitha, Prajna, Sivakumar Rathinam, Krishnadas Banushree, Shanmugam Maheshkumar, Rajendran Vijayakumar, and Padmakar Sathe. "Ocular involvement associated with an epidemic outbreak of chikungunya virus infection." American journal of ophthalmology 144, no. 4 (2007): 552-556..
23. Laskowski, Roman A., Malcolm W. MacArthur, David S. Moss, and Janet M. Thornton. "PROCHECK: a program to check the stereochemical quality of protein structures." Journal of applied crystallography 26, no. 2 (1993): 283-291.
24. Laskowski, Roman A. "Enhancing the functional annotation of PDB structures in PDBsum using key figures extracted from the literature." Bioinformatics 23, no. 14 (2007): 1824-1827.
25. Laurent, Philippe, Karin Le Roux, Philippe Grivard, Gérard Bertil, Florence Naze, Miguel Picard, Frédérik Staikowsky, Georges Barau, Isabelle Schuffenecker, and Alain Michault. "Development of a sensitive real-time reverse transcriptase PCR assay with an internal control to detect and quantify chikungunya virus." Clinical chemistry 53, no. 8 (2007): 1408-1414.
26. LEHTOVAARA, Päivi, Ismo ULMANEN, Leevi KÄÄRIÄINEN, Sirkka KERÄNEN, and Lennart PHILIPSON. "Synthesis and Processing of Semliki Forest Virus-Specific Nonstructural Proteins in vivo and in vitro." European Journal of Biochemistry 112, no. 3 (1980): 461-468.
27. Levinson, Randy S., James H. Strauss, and Ellen G. Strauss. "Complete sequence of the genomic RNA of O'nyong-nyong virus and its use in the construction of alphavirus phylogenetic trees." Virology 175, no. 1 (1990): 110-123.
28. Litzba, Nadine, Isabelle Schuffenecker, Hervé Zeller, Christian Drosten, Petra Emmerich, Remi Charrel, Petra Kreher, and Matthias Niedrig. "Evaluation of the first commercial chikungunya virus indirect immunofluorescence test." Journal of virological methods 149, no. 1 (2008): 175-179.
29. Lumsden, W. H. R. "An epidemic of virus disease in Southern Province, Tanganyika territory, in 1952–1953 II. General description and epidemiology."Transactions of the Royal Society of Tropical Medicine and Hygiene 49, no. 1 (1955): 33-57.
30. Mahendradas, Padmamalini, Shylaja K. Ranganna, Rohit Shetty, Ramgopal Balu, Kannan M. Narayana, Rajesh B. Babu, and Bhujang K. Shetty. "Ocular manifestations associated with chikungunya." Ophthalmology 115, no. 2 (2008): 287-291.

31. Maiti, C. R., A. K. Mukherjee, B. Bose, and G. L. Saha. "Myopericarditis following chikungunya virus infection." Journal of the Indian Medical Association70, no. 11 (1978): 256-258.
32. Mittal, Apoorva, Saurabh Mittal, Jayahar M. Bharathi, R. Ramakrishnan, and Padmakar S. Sathe. "Uveitis during outbreak of Chikungunya fever."Ophthalmology 114, no. 9 (2007): 1798-1798.
33. Mourya, D. T., and A. C. Mishra. "Chikungunya fever." The Lancet 368, no. 9531 (2006): 186-187.
34. Murshudov, Garib N., Alexei A. Vagin, and Eleanor J. Dodson. "Refinement of macromolecular structures by the maximum-likelihood method." Acta Crystallographica Section D: Biological Crystallography 53, no. 3 (1997): 240-255.
35. Myers, Ruth M., and Donald E. Carey. "Concurrent isolation from patient of two arboviruses, chikungunya and dengue type 2." Science 157, no. 3794 (1967): 1307-1308.
36. Nimmannitya, Suchitra, Scott B. Halstead, Sanford N. Cohen, and Mark R. Margiotta. "Dengue and Chikungunya Virus Infection in Man in Thailand, 1962–1964 I. Observations on Hospitalized Patients with Hemorrhagic Fever." The American journal of tropical medicine and hygiene 18, no. 6 (1969): 954-971.
37. Obeyesekere, Ivor, and Yvette Hermon. "Myocarditis and cardiomyopathy after arbovirus infections (dengue and chikungunya fever)." British heart journal 34, no. 8 (1972): 821.
38. Otwinowski, Zbyszek, Wladek Minor, and Charles Carter W Jr. "Processing of X-ray diffraction data collected in oscillation mode." (1997): 307-326.
39. Padbidri, V. S., and T. T. Gnaneswar. "Epidemiological investigations of chikungunya epidemic at Barsi, Maharashtra state, India." Journal of hygiene, epidemiology, microbiology, and immunology 23, no. 4 (1978): 445-451.
40. Panning, Marcus, Klaus Grywna, Marjan Van Esbroeck, Petra Emmerich, and Christian Drosten. "Chikungunya fever in travelers returning to Europe from the Indian Ocean region, 2006." Emerging infectious diseases 14, no. 3 (2008): 416.
41. Peränen, Johan, Pirjo Laakkonen, Marko Hyvönen, and Leevi Kääriäinen. "The alphavirus replicase protein nsP1 is membrane-associated and has affinity to endocytic organelles." Virology 208, no. 2 (1995): 610-620.

42. Powers, Ann M., Aaron C. Brault, Yukio Shirako, Ellen G. Strauss, WenLi Kang, James H. Strauss, and Scott C. Weaver. "Evolutionary relationships and systematics of the alphaviruses." Journal of Virology 75, no. 21 (2001): 10118-10131.
43. Powers, Ann M., and Christopher H. Logue. "Changing patterns of chikungunya virus: re-emergence of a zoonotic arbovirus." Journal of General Virology 88, no. 9 (2007): 2363-2377.
44. Rawlings, N. D., D. P. Tolle, and A. J. Barrett. "MEROPS: the peptidase database Nucleic Acids Res,. 32." D160 D164 (2004).
45. Rawlings, Neil D., and Alan J. Barrett. "Families of cysteine peptidases."Methods in enzymology 244 (1993): 461-486.
46. Reiter, Paul, Didier Fontenille, and Christophe Paupy. "Aedes albopictus as an epidemic vector of chikungunya virus: another emerging problem?." The Lancet infectious diseases 6, no. 8 (2006): 463-464.
47. Rezza, G., L. Nicoletti, R. Angelini, R. Romi, A. C. Finarelli, M. Panning, P. Cordioli et al. "Infection with chikungunya virus in Italy: an outbreak in a temperate region." The Lancet 370, no. 9602 (2007): 1840-1846.
48. Robinson, Marion C. "An epidemic of virus disease in Southern Province, Tanganyika territory, in 1952–1953." Transactions of the Royal Society of Tropical Medicine and Hygiene 49, no. 1 (1955): 28-32.
49. Salonen, Anne, Lidia Vasiljeva, Andres Merits, Julia Magden, Eija Jokitalo, and Leevi Kääriäinen. "Properly folded nonstructural polyprotein directs the Semliki Forest virus replication complex to the endosomal compartment." Journal of virology 77, no. 3 (2003): 1691-1702.
50. Schwartz, Olivier, and Matthew L. Albert. "Biology and pathogenesis of chikungunya virus." Nature Reviews Microbiology 8, no. 7 (2010): 491-500.
51. Sergon, Kibet, Charles Njuguna, Rosalia Kalani, Victor Ofula, Clayton Onyango, Limbaso S. Konongoi, Sheryl Bedno et al. "Seroprevalence of chikungunya virus (CHIKV) infection on Lamu Island, Kenya, October 2004." The American journal of tropical medicine and hygiene 78, no. 2 (2008): 333-337.
52. Sergon, Kibet, Ali Ahmed Yahaya, Jennifer Brown, Said A. Bedja, Mohammed Mlindasse, Naphtali Agata, Yokouide Allaranger et al. "Seroprevalence of Chikungunya virus infection on

Grande Comore Island, union of the Comoros, 2005." The American journal of tropical medicine and hygiene 76, no. 6 (2007): 1189-1193.

53. Shah, K. V., C. J. Gibbs Jr, and G. Banerjee. "Virological Investigation of the Epidemic of Haemorrhagic Fever in Calcutta: Isolation of Three Strains of Chikungunya Virus." The Indian journal of medical research 52 (1964): 676-683.
54. Simizu, B., K. Yamamoto, K. Hashimoto, and T. Ogata. "Structural proteins of Chikungunya virus." Journal of virology 51, no. 1 (1984): 254-258.
55. Sissoko, Daouda, Denis Malvy, Khaled Ezzedine, Philippe Renault, Frederic Moscetti, Martine Ledrans, and Vincent Pierre. "Post-epidemic Chikungunya disease on Reunion Island: course of rheumatic manifestations and associated factors over a 15-month period." PLoS Negl Trop Dis 3, no. 3 (2009): e389.
56. Simon, Fabrice, Philippe Parola, Marc Grandadam, Sabrina Fourcade, Manuela Oliver, Philippe Brouqui, Pierre Hance et al. "Chikungunya infection: an emerging rheumatism among travelers returned from Indian Ocean islands. Report of 47 cases." Medicine 86, no. 3 (2007): 123-137.
57. Staples, J. Erin, Robert F. Breiman, and Ann M. Powers. "Chikungunya fever: an epidemiological review of a re-emerging infectious disease." Clinical Infectious Diseases 49, no. 6 (2009): 942-948.
58. Strauss, Ellen G., Raoul J. De Groot, Randy Levinson, and James H. Strauss. "Identification of the active site residues in the nsP2 proteinase of Sindbis virus." Virology 191, no. 2 (1992): 932-940.
59. Tsetsarkin, Konstantin A., Dana L. Vanlandingham, Charles E. McGee, and Stephen Higgs. "A single mutation in chikungunya virus affects vector specificity and epidemic potential." PLoS Pathog 3, no. 12 (2007): e201.
60. Vagin, Alexei, and Alexei Teplyakov. "MOLREP: an automated program for molecular replacement." Journal of applied crystallography 30, no. 6 (1997): 1022-1025.
61. Vasiljeva, Lidia, Leena Valmu, Leevi Kääriäinen, and Andres Merits. "Site-specific protease activity of the carboxyl-terminal domain of Semliki Forest virus replicase protein nsP2." Journal of Biological Chemistry 276, no. 33 (2001): 30786-30793.
62. Vasiljeva, Lidia, Andres Merits, Andrey Golubtsov, Valeria Sizemskaja, Leevi Kääriäinen, and Tero Ahola. "Regulation of the sequential processing of Semliki Forest virus replicase polyprotein." Journal of Biological Chemistry 278, no. 43 (2003): 41636-41645.

63. Vazeille, Marie, Sara Moutailler, Daniel Coudrier, Claudine Rousseaux, Huot Khun, Michel Huerre, Julien Thiria et al. "Two Chikungunya isolates from the outbreak of La Reunion (Indian Ocean) exhibit different patterns of infection in the mosquito, Aedes albopictus." PloS one 2, no. 11 (2007): e1168.
64. Weaver, Scott C., and William K. Reisen. "Present and future arboviral threats."Antiviral research 85, no. 2 (2010): 328-345.
65. Yazdani, R., and V. V. Kaushik. "Chikungunya fever." Rheumatology 46, no. 7 (2007): 1214-1215.

Chapter III:

AURAV nsP2 papain-like protease cloning, expression, purification and characterization.

3.1 Introduction

Aura virus is an alphavirus member which was first isolated from *Aedes serratus* mosquito in north Argentina and Brazil [2] and belongs to Sindbis virus (SINV) lineage [1]. Geographical distribution of Aura virus is wide, still no case of human infection and no information regarding its vertebrate host is known [3]. Aura is serologically related to both WEEV and SINV lineage but is still having some differences from both the lineages [4]. Some structural aspects of Aura virus are similar to SINV which are observed to be different from SFV/RRV lineage, like glycoprotein spikes which have close resemblance with SINV spikes [3]. WEEV is showed to be a recombinant of EEEV and SINV deriving maximum part of its genome from the former while part of 3' NTR region and two glycoproteins are derived from the latter virus lineage [5, 6]. All the SINV lineage members have some conserved signatures like 3'NTR's, which is the hallmark of this lineage. This signature element is variable with other alphavirus lineage except EEEV [13, 14]. Sequence similarity between Aura and Sindbis is not much compared to what is observed from other members. However it is found to be more related to SINV lineage than WEEV [3].

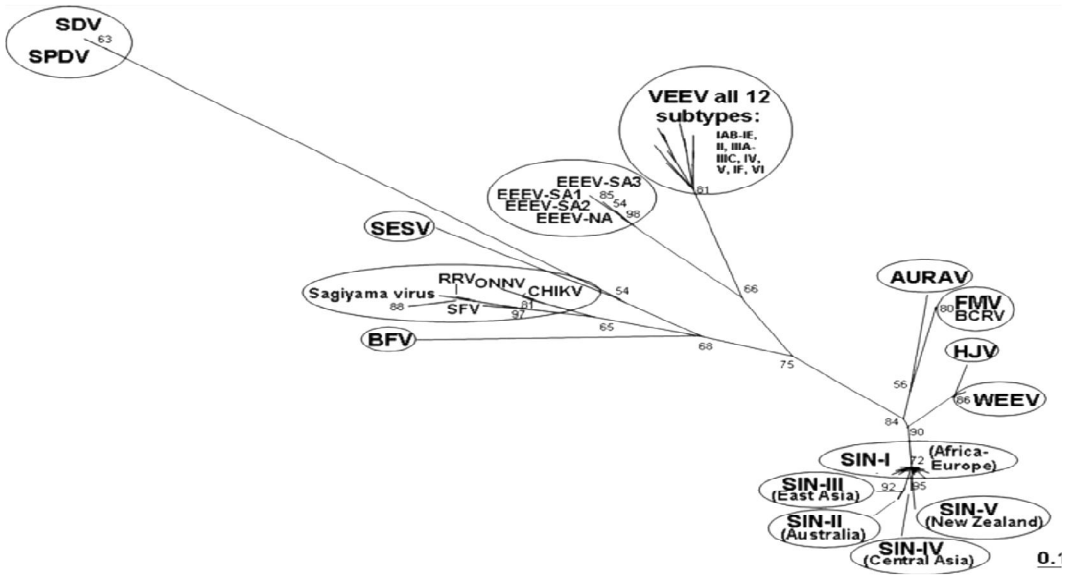


Figure 3.1 Radiogram of deduced amino acid sequences for residues 117–229 of the E2 region of SIN complex viruses, recombinant alphaviruses, Aura virus, and reference strains of all known alphavirus species as obtained after applying the Dayhoff model of substitutions implemented in PUZZLE. Bootstrap values are indicated at the main

bifurcations separating the alphavirus species and respective genotypes. Branch lengths are proportional to genetic relatedness of the respective Alphavirus species and genotypes [1].

It is known that alphaviruses initially originated in New World and then spread out (twice) with evolution to the Old world, first to form SINV lineage and second time to form SFV lineage [7]. Even though Aura is similar to SINV it has diverged significantly by sharing 73% similarity in non-structural proteins and 62% in structural proteins. Such case of diversion was also observed in case of EEV and VEEV, which has major changes in 3'NTR region and structural proteins making these two almost similar viruses members of different sub-groups [3]. With the careful sequence observations and lineage studies it has been concluded that around 1000-2000 years ago, SINV initially a New world member diverged to form two different lineages one is Aura virus and rest other consisting of Sindbis and Sindbis-like viruses [13, 16]. And again after some time this new Sindbis lineage shifted to Old world territory and with a new event developed WEEV after undergoing a recombination event with EEV [15, 16]. So, in conclusion it can be said that Aura although is a member of SINV lineage but it acts as a representation of New World in the SINV lineage.

Thus, studying nsP2 (non structural protein 2) from this virus might hint us about the changes which the virus would have undergone so as to become unique and one of a kind in its lineage. For this purpose we cloned, expressed and purified the C terminal nsP2 protease (472-800)from AURA virus in a prokaryotic expression system. And also developed a simple β-galactosidase based assay for characterizing its proteolytic activity.

3.2 Cloning, expression and purification of nsP2Aura protease (472-800 amino acid).

3.2.1 Materials

3.2.1.1 Chemicals

Chemicals and reagents used for running and purification stocks were of ACS grade and purchased either from Himedia, Merck or Sigma. Isopropyl-D-β-galactopyranoside (IPTG) was purchased from Himedia chemicals. Crystallization solutions were purchased from Hampton research and Molecular Dimensions. Hitrap SP HP 5ml columns from GE Healthcare, superdex 75 16/60 column from GE Healthcare, General chemicals from Merck India and Sigma Aldrich.

3.2.1.2 Molecular biology Chemicals

Restriction enzymes like *Nco*I, *Bam*HI, *Xho*I and *Nde*I were purchased along with their reaction buffers from New England Biolabs. T4 DNA ligase, DNaseI, RNase A, *Pfu* polymerase and Taq polymerase were also purchased from New England Biolabs.

3.2.1.3 Bacterial expression vector and strains

Escherichia coli DH5α and *Rossetta* (DE3) strains for cloning and expression respectively were purchased from Novagen (USA). pSV plasmid for β-galactosidase was gift from Dr. Partha Roy, IIT roorkee.

3.2.1.4 Oligonucleotides for PCR amplification

The oligonucleotides for the PCR amplification of nsP2AURA gene, β-galactosidase amplification and addition of peptide sequence between the β-galactosidase were designed with the help of oligoanalyser server of IDT technologies. Different restriction sites were added in the forward and reverse primers (Table no. 3.1).

3.2.1.5 Culture media and antibiotics

LB broth and LB agar media for culture was purchased from Merck speciality chemicals as dried granules. The Luria-Bertani (LB) culture medium: 1% (w/v) tryptone, 0.5% yeast extract, 1% (w/v) NaCl was sterilized by autoclaving. Antibiotics for selective growth of the cells

kanamycin (stock solution: 100 mg/mL in ddH2O, working concentration 50 µg/mL) chloramphenicol (stock solution: 100 mg/mL, in absolute ethanol, working concentration: 35 µg/mL) were purchased from Himedia Chemicals.

Table 3.1 Oligonucleotides for the cloning of nsP2AURA and recombinant β-galactosidase

Gene name	Insertion Vector	Primer direction	Primer sequence	Incorporated restriction enzyme site
nsP2AURA (472-800)	pET28c	F	GATTCTCCATGGGACCCTTTCGCCA GCAAAGTG	*NcoI*
		R	ATTCTCTCGAGTTAATAATTGTCGAA TATATTGGATACTGC	*XhoI*
β-galactosidase^{R} (Part I Primers)	pET28c	F	GATTCTCATATGGTCGTTTTACAACG TCGTGACT	*NdeI*
		R	GATTCTGGATCCGTTTTCATCATATT TAATCAGCGACTG	*BamHI*
β-galactosidase^{R} (nsP2AURA 1/2 site)	pET28c	F1	CTAGTAGAAACTCCAGGCAACCCTG GTCGGCTTAC	No RE
		F2	GGAGCTGCTCTAGTAGAAACTCCA G GCAACCCTTGG	No RE
		F3	GATTCTGGATCCGTTGATGATGCT GAGCTGCTCTAGTA	*BamHI*
β-galactosidase^{R} (nsP2AURA 3/4 site)	pET28c	F1	ATATTTTCTACAGACGGCAACCCATG GTCGGCTTAC	No RE
		F2	GGTGTAGGTGGGTACATATTTTCTAC AGACGGCAACCCA	No RE
		F3	GATTCTGGATCCCTAACCGGTGTAG GTGGGTACATATTT	*BamHI*
		R	GATTCTCTCGAGTTATTTTTGACTCC AGACCAACTCGTA	*XhoI*
β-galactosidase^{R} (nsP2AURA 2/3 site)	pET28c	F1	CCTTCTTATCGTGTAGGCAACCCTTG GTCG GCTTAC	No RE
		F2	TCAGGTGCAGCTCCTTCTTATCGTGT AGGCAACCCT	No RE
		F3	GATTCTGGATCCAAAGATGGTTCAG GTGCAGCTCCTTCT	*BamHI*

3.2.1.6 Crystallisation Solutions

Crystallization screens (Crystal Screen I & II, PEG/ion I & II, Index, Salt and Crystal Screen Cryo) were procured from Hampton Research. For optimization reagents were made by highest purity ACS grade chemicals. To remove insoluble particles, solutions were filtered through 0.22 μ filter. The prepared reagents were maintained 4°C. PEG solutions were prepared by overnight stirring and stored in light protected bottles.

3.2.2 Methods

3.2.2.1 Cloning of nsP2AURA gene in pET-28c vector

The full length region coding for nsP2AURA protease was PCR amplified using forward and reverse primers containing *Nco*I and *Xho*I restriction sites (See materials). The amplified fragment was double digested with both the restriction enzymes and ligated into pET-28c vector in between *Nco*I and *Xho*I restriction sites without 6xhis tag at N-terminal. The pET-28c plasmid ligated with nsP2AURA was transformed into *E. coli* DH5α cells. The cloned plasmid was isolated from *E. coli* DH5α cells and purified by DNA clean and concentrate kit from Zymo research, USA and clone was confirmed by PCR (Figure3.2 A), restriction digestion (Figure 3.2 B) and automated DNA sequencing with internal primers (Ocimum Biosolutions, Bangalore, India).

3.2.2.2 Expression optimization of nsP2AURA

After the clone confirmation, the pET-28c-nsP2AURA plasmid was transformed into *E. coli Rossetta* (DE3) expression host. The expression analysis and optimization of new constructs were performed with small scale culture. Single colony from the transformed plate was used to inoculate 10 mL of LB broth containing 50 μg/mL of kanamycin in a 50 mL glass culture tube. The 100 μL of this overnight grown culture was used to inoculate fresh 10 mL LB broth containing 50 μg/ml kanamycin. The inoculated culture was allow to grow at 37°C at 200 rpm shaking until the optical density at 600 nm reached to 0.6-0.7. The culture was then induced with different concentrations of IPTG (0.25mM to 1.0 mM) and then transferred to 18°C, 25°C and 37°C temperatures and was allowed to grow for 16 hrs, 6 hrs and 3 hrs respectively. Then the cells were harvested by centrifugation in a cooling micro centrifuge at 3214 g for 15 min. and stored on ice. The collected cell pellet was resuspended in buffer A (50 mM phosphate pH 7.5,

5% glycerol, 0.5 mM DTT) and lysed using an ultra sonicator. To remove the cell debris, lysate was centrifuged at 28,928 g for 20 min. at 4°C. The expression profile was analyzed by running the supernatant and pellet on 12% SDS-PAGE (Figure 3.3).

3.2.2.3 Purification of nsP2AURA

Two litre culture of *Rossetta* (DE3) cells, containing pET-28c-nsP2AURA were grown for 20 hrs at 18°C. Each litre of cells was grown in a 2 litre baffled flask at 200 rpm. Following growth cells were pellet down by centrifugation at (3214 g, 10min.) and resuspended in 30 mL of buffer A (50 mM phosphate pH 7.5, 5% Glycerol, 0.5 mM DTT) containing 0.5 mg/mL of lysozyme and 0.01 mg/mL of DNaseI (with 10 mM MgCl2). The cells were incubated on ice for 10 min. and then lysed using French press (Constant cell Disruptor Systems, UK). The disrupted cells were then subjected to centrifugation so as to separate the soluble portion from the insoluble fraction (at 28,928 g, 60 min., 4°C). In the meantime the HiTrap SP HP column was equilibrated with the buffer A. The cleared cell lysate was then loaded on the equilibrated column at the flow rate of 0.4 mL/min. Protein was eluted with 10x column volume linear gradient from 0% to 100% buffer B (50 mM phosphate pH 7.5, 5% glycerol, 1M NaCl, 0.5 mM DTT), at the rate of 2 mL/min followed by isocratic wash with 100% Buffer B. 5 mL fractions were collected over the course of gradient. Sample from fractions were run on 12% SDS-PAGE gel (Figure 3.4 B). Fractions containing the nsP2AURA were pooled (based on the absorbance peak and SDS-PAGE analysis) and dialysed against the dialysis buffer I (50mM phosphate pH 7.5, 5% Glycerol, 0.5 mM DTT). The dialysed protein was then concentrated using Amicon (10 kDa cut-off) till 1 mL final volume and loaded on pre-equilibrated superdex 75 16/60 (50mM phosphate pH 7.5, 5% glycerol, 0.5 mM DTT) column. Protein eluted at the correct column volume was then concentrated using Amicon (10 kDa cut-off), only after confirmation by loading on the SDS-PAGE (Figure 3.4 D). Protein was concentrated to around 10 mg/mL.

3.2.2.4 Determination of protein concentration

Bradford method [12] was utilized to estimate the nsP2AURA protein concentration using Bradford reagent provided by Bio-Rad Laboratories, USA. Supplier suggested protocol was followed for the assay. Briefly 5 µL of protein sample was mixed with 795 µL of buffer used for purification. 200 µL of Bradford's reagent was added to it and incubated for 1 min. The absorbance reading at 595 nm wavelength was determined using Perkin Elmer Lambda 25 UV-

visible spectrophotometer. The protein concentration was calculated by interpolating the absorbance reading into a standard BSA curve.

3.2.2.5 Crystallisation

Crystallization trials of nsP2AURA protease were done by sitting drop and hanging drop vapor diffusion method. The purified protein was concentrated to 10 mg/mL and buffer exchanged with 25 mM Tris buffer (pH 7.5), 5% glycerol and 100 mM NaCl. 1 µL of the purified protein solution was mixed with 1 µL of reservoir solution containing different combination of precipitants, buffers and salts. The drop was equilibrated at 4°C and 20°C. However, the crystallization trials did not produce nsP2AURA crystals. Therefore, for structural analysis we built *in silico* homology model of the protease domain from Aura virus.

3.3 *In Silico* Structure modeling

The homology model of nsP2AURA (472-800 amino acid) was generated by the following steps: template selection (based on the similarity) from Protein Data Bank (PDB), sequence-template alignment, model building, model refinement and validation. The template for homology modeling was searched using NCBI BLAST [27] search tool against PDB database. BlastP program was run with default values having BLOSUM62 as a scoring matrix, word size of 3, gap penalty of 11 and gap extension penalty of 1. Crystal structure of nsP23 SINV (PDB ID: 4GUA) was obtained as the best hit in BlastP search which shows 58% sequence identity with nsP2AURA. Thus, nsP23 SINV crystal structure (4GUA) was used as template to generate a comparative 3D model of nsP2AURA using MODELLER *9v13* [9]. The multiple sequence alignment of query sequence with template sequence was performed using ClustalW program [17]. The alignment file was used as an input in PIR/PAR format for model generation in MODELLER [9]. Several preliminary models were generated using MODELLER which were ranked on the basis of their DOPE scores. Five sets of models were selected on the basis of lowest DOPE scores and stereo-chemical quality of each was evaluated using PROCHECK. The models having least number of residues in the disallowed region were further refined for relieving steric clashes and improper contacts. Energy minimization of the top screened model was performed using Swiss-Pdb Viewer 4.01 (http://www.expasy.org/spdbv/) that implements GROMOS96 force field to compute energy and to execute energy minimization. PROCHECK

was again used to inspect the stereo-chemical quality of the model. Structural validation was evaluated using ERRAT plot of SAVES server that gives a measure of the structural error at each residue in a generated protein model (http:// nihserver.mbi.ucla.edu/SAVES/). This process was repeated iteratively till most of the amino acid residues attained a cut-off value below 95% in ERRAT plot. The refined model was further validated by VERIFY-3D of SAVES server [10].

3.4 β-galactosidase based activity assay

3.4.1 Material

pSV plasmid which contains β-galactosidase gene from amino acid 1-1024 was used as the initial template source of the β-galactosidase. The oligonucleotides used for adding the nsP2AURA different cleavage sites were described in section 3.3.4. The *Nde*I, *Bam*HI and *Xho*I restriction enzymes were from New England Biolabs. ONPG (o-nitrophenyl β-D-galactopyranoside) was purchased from Sigma Chemicals. Cary UV plate reader was used for absorbance at 420 nm. Z buffer (0.06 M di-sodium hydrogen phosphate, 0.04 M sodium dihydrogen phosphate, 0.01 M potassium chloride and 0.001 M magnesium sulphate) for activity.

3.4.2 Methods

3.4.2.1 Cloning of the β-galactosidase gene with the nsP1/2, nsP2/3 and nsP3/4 sites

Cloning for developing β-galactosidase^{R} (modified β- galactosidase) gene was done in two steps. First step includes cloning of the 1-578 amino acid region of β-galactosidase gene by PCR amplification of the gene. The amplified gene was product was PCR purified using Qiagen PCR purification kit and then double digested with *Nde*I and *Bam*HI enzymes at 37 °C for 60 min. The pET-28c plasmid was also double digested with the *Nde*I and *Bam*HI enzyme. Both the digested PCR product and the linearized pET-28c plasmid were gel extracted and then ligated in the presence of T4 DNA ligase at 16 °C for 15 hrs. The ligation mixture was directly used for the transform into $CaCl_2$ competent DH5α cells by heat shock method [11]. Transformed DH5α cells were screened by kanamycin resistant. Individual colonies were picked, grown overnight in 5 mL LB broth at 37 °C with shaking. Further, plasmids were isolated from overnight growth using commercial mini-prep kit (Qiagen, Inc. Valencia, CA). Restriction digestion of the

isolated plasmids was done to select the construct containing the correct size insert. After this second part was amplified; using combination of different forward primers the peptide cleavage site was added at the 5' terminus of the second part of the β-galactosidase gene. The amplified part was cloned in the pET-28c-β-galactosidase (1-578) vector by using the above methodology. The same method was used for cloning the β-galactosidase gene with nsP2AURA 1/2, 2/3 and 3/4 cleavage sites.

3.4.2.2 Purification of β-galactosidase^{R} by Affinity Chromatography

One liter culture of *BL21* (DE3) cells, containing pET-28c- β-galactosidase^{R} were grown for 4 hrs at 37°C. Each litre of cells was grown in a 2 litre baffled flask at 200 rpm. Following growth cells were pellet down by centrifugation at (3214 g, 10 min.) and resuspended in 30 mL of buffer A (50 mM phosphate pH 7.5, 5% Glycerol, 500 mM NaCl) containing 0.5 mg/mL of lysozyme and 0.01 mg/mL of DNaseI (with 10 mM MgCl2). The cells were incubated for 10 min. and then lysed using French press (Constant cell Disruptor Systems, UK). The disrupted cells were then subjected to centrifugation so as to separate the soluble portion from the insoluble fraction (at 28,928 g, 60min., 4°C). In the meantime the His trap HP column was equilibrated with the buffer A. The cleared cell lysate was then loaded on the equilibrated column at the flow rate of 0.4 mL/min. Protein was eluted with 10x column volume linear gradient from 0% to 100% buffer B (50 mM phosphate pH 7.5, 5% glycerol, 250 mM NaCl), at the rate of 2 mL/min followed by isocratic wash with 100% buffer B. 5 mL fractions were collected over the course of gradient and isocratic wash. Fractions containing the β-galactosidase^{R} were pooled (based on the absorbance peak and 12% SDS-PAGE analysis) and dialysed against the dialysis buffer I (50 mM phosphate pH 7.5, 5% glycerol). The dialysed protein was then concentrated using Amicon (10kDa cut-off), only after confirmation by loading on the SDS-PAGE. Protein was concentrated to around 20 mg/mL. Same methodology was used for purification of all the β-galactosidase^{R} proteins containing nsP2AURA 1/2, 2/3 and 3/4 sites.

3.4.2.3 B-galactosidase based activity assay

There were two phases of the assay; (i) Validating the activity by Gel electrophoresis and (ii) In microtitre plate to optimize the standard assay conditions.

The purified β-galactosidase^{R} was incubated with the purified nsP2AURA in 2:1 molar concentrations at different temperature conditions. The incubated mixtures were mixed by 6x loading buffer and loaded on 10% SDS-PAGE gel. The gel was stained and destained using the standard procedure.

(ii) The experiment was done in triplicate by repeating the above conditions of the reaction in the microtitre plate. The assay was performed by incubating the different molar concentrations of pET-28c-β-galactosidase^{R} protein in the Z buffer. After that ONPG [o-nitrophenyl β-D-galactopyranoside] with final concentration of 10 µM was added to the well and the mix was incubated at 37 °C for 30 min. The reaction was halt by the addition of 100 µL of 1M calcium carbonate. The readings were taken at 420 nm. Then the experiment was performed by fixing the substrate i.e β-galactosidase^{R} concentration with varying concentration of nsP2AURA. The reaction was performed in the presence of Z buffer at 37 °C for 30 min. in triplicates. The After that ONPG with final concentration of 10µM was added to the well and the mix was incubated at 37 °C for 30 min. Reaction was stopped after the addition of 100ul of 1M calcium carbonate, absorbance was taken at 420nm.

3.5 Results and Discussion

3.5.1 Cloning and expression optimization of nsP2AURA protease

pET-28c-nsP2AURA (472-800 amino acid) was cloned in the pET-28c vector between *Nco*I and *Xho*I sites, without any tag to skip the tag cleavage step. The viral protein sequence of 472-800 of nsP2 protein, shown to contain the proteolytic domain, was cloned in the plasmid pET-28c.The cDNA of AURA was isolated and used for the PCR amplification of nsP2AURA gene. Analysis of PCR product (Figure 3.2 A) on 1% agarose gel electrophoresis confirmed the presence of a single amplification band of approx. 1000 bp size, corresponding to the nsP2 gene size (987 bp). The forward and reverse primers with restriction endonuclease sites of *NcoI* and *XhoI* were used for the PCR amplification. The vector DNA and PCR product were double digested with *NcoI* and *XhoI* restriction enzymes and T4 DNA ligase was used to ligate digested PCR product into digested vector. The ligated product was successfully transformed in to *E. coli* DH5α cells. To confirm the cloning of gene into the vector, plasmid DNA isolated from the transformed cells and subjected to double digestion with *Nco*I and *Xho*I yielded the fragment of

desired nsP2AURA gene length (987 bp) (Figure 3.2 B). The automated DNA sequencing further confirmed the specificity of cloned DNA fragment. In order to optimize the nsP2AURA expression, the recombinant vector was transformed in to *E. coli Rossetta* (DE3) cells. Small scale expression of nsP2AURA was analyzed at five different IPTG concentrations (0.1 mM to 0.5 mM) and three different temperatures (18°C, 25°C and 37°C). The optimum expression and solubility was observed with 0.4 mM IPTG concentration at 25°C temperature (Figure 3.3). Cloned in pET-28c between *Nco*I and *Xho*I sites, the protein is expressed as a native, non-tagged protein and thus purification was done by the ion exchange chromatography.

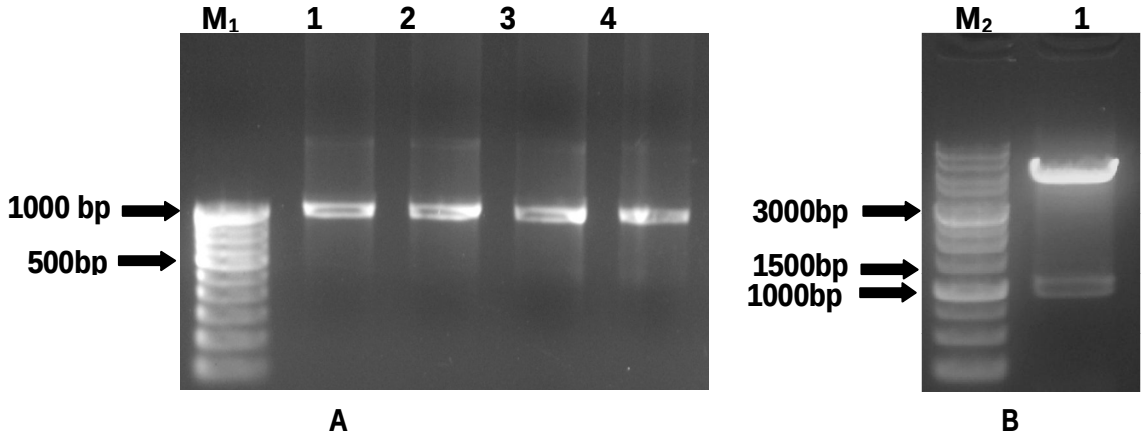


Figure 3.2 Amplification and cloning of nsP2AURA (472-800) gene. **Agarose gel electrophoresis (A)** PCR products. lane M1: DNA marker; lane 1-4: PCR product amplified from AURA cDNA with nsP2AURA gene specific primers. **(B)** Recombinant plasmids digested with *NcoI* and *XhoI*. lane M2: DNA marker; lane 1: pET-28c- nsP2AURA from the screened colonies.

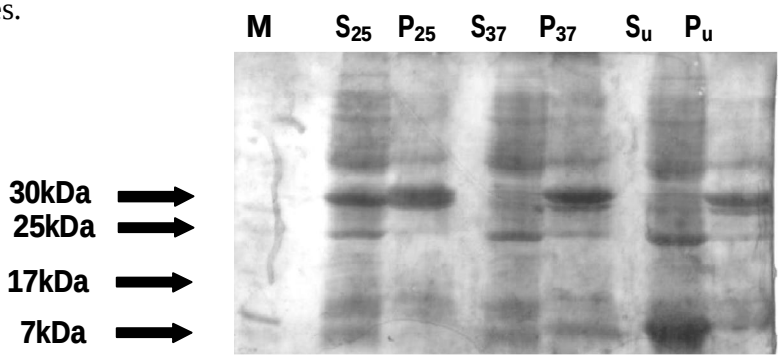
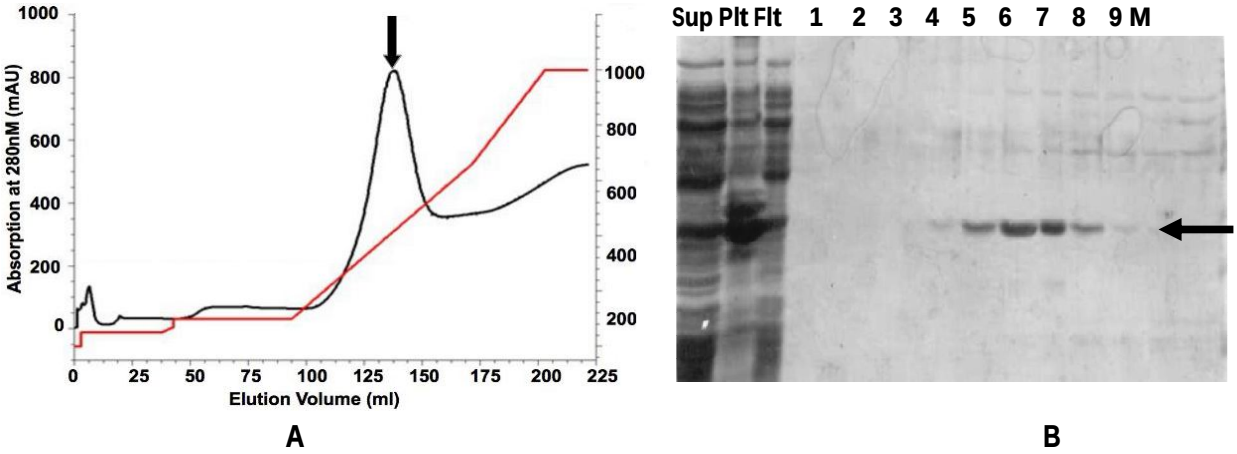


Figure 3.3 Expression optimization of nsP2AURA gene. 12% SDS-PAGE analysis. Lane M: Protein ladder; P_{25} – Pellet at 25°C, P_{37} – Pellet at 37°C, P_u - Uninduced Pellet, S_{25} - Supernatant 25°C, S_{37} - Supernatant 37°C, S_u - Uninduced Supernatant.

3.5.2 Purification and Crystallization of nsP2AURA

The plasmid was transformed in *Rossetta* (DE3) cells for expression purpose, which were induced with 0.4 mM IPTG at 25 °C for 6 hrs. Purification of nsP2AURA protein was pursued from 2 litres of cell pellets. Cells were lysed after gentle treatment with lysozyme followed by French press. Protease inhibitor cocktails were not used to avoid any effect on the protease activity of the enzyme, and hence to maintain the stability of the protein all the process were performed at 4°C. After disruption, high speed centrifuge with fixed angle rotor was used to separate the debris from the soluble fraction containing the protein. The calculated pI for the nsP2AURA was found to be 9.3, indicating it would be positively charged at pH 8.0 and thus should bind to a cation exchange column. After this, the soluble fraction was loaded on a Hi trap SP column and eluted with a NaCl gradient (Figure 3.4 A). Protein obtained after this step was impure (Figure 3.4 B), therefore to get >90% pure protein, in the next step size exclusion chromatography was done. The monomeric state of the nsP2 protein in case of the other viruses has been reported [8], so use of superdex 75 16/60 column was considered appropriate for gel-filtration step. The protein fractions eluted from the gel filtration column were loaded on the 12% SDS-PAGE gel and were pure (Figure 3.4 C and D). Commercial sparse matrix screens from Hampton research, USA were utilized for the initial crystal screening. Each screen had 96 different solutions varying in precipitant, salt and additives combination. Different buffers were used to screen a wide pH range (4.0-10.0). The concentration of nsP2AURA was kept at 10 mg/mL for the initial screening.



A B

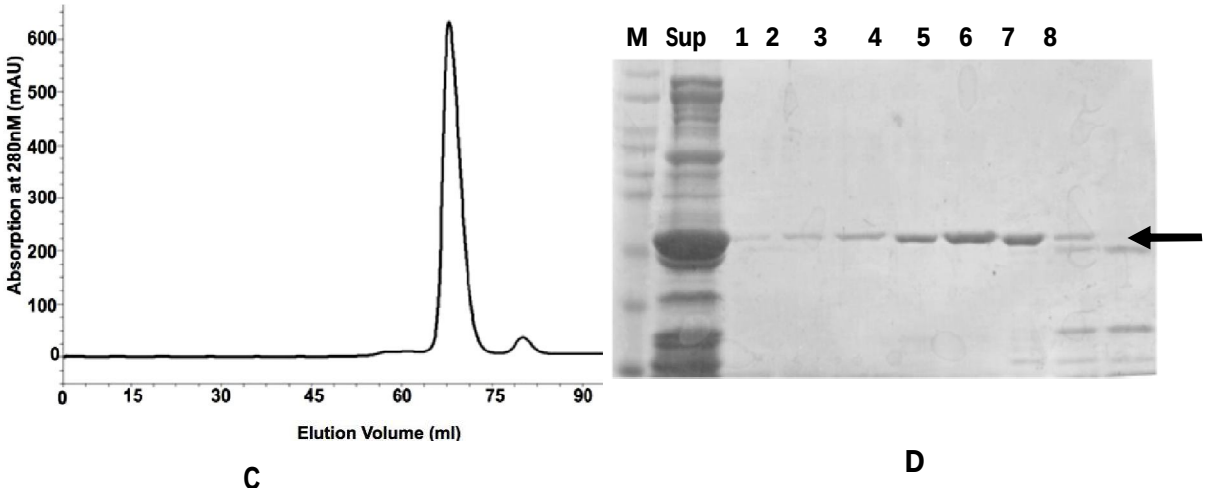


Figure 3.4 Optimized purification of recombinant nsP2AURA. (**A)** Chromatograms obtained from an extract of culture of *E. coli Rossetta* (DE3) cells containing the pET-28c- nsP2AURA applied to a cation exchange chromatography column with peak corresponding to fractions containing nsP2AURA indicated by an arrow, **(B)** Analysis of the cation exchange purification by SDS-PAGE (12%) stained with Comassie blue. Soluble extract of E. coli culture (Sup), insoluble extract of *E.coli* culture (Pel), flow through from the Hi trap column (FLT) and eluted fractions (lanes 1-9), protein of 36 kDa marked by an arrow**.(C)** and from the fraction purified by ion exchange applied to a superdex-75 gel filtration column. (**D**) Protein fractions from gel filtration chromatography. M is the protein ladder. 36 kDa nsP2AURA marked by an arrow.

3.5.3 Structure analysis of *in-silico* homology model

nsP2AURA shares 58 % identity and 63 % similarity (with template sequence (i.e. nsP23 SINV). The predicted secondary structural elements of nsP2AURA are roughly identical to the secondary structures of nsP2SINV (Figure 3.5). nsP2AURA homology model was constructed on the basis of crystallographic structure of nsP23 SINV (PDB ID: 4GUA) [21]. The generated model was subjected to refinement and energy minimization. PROCHECK, Verify-3D and ERRAT plot [10] were used to evaluate the stereo-chemical parameters of the energy minimized model of nsP2AURA. Ramachandran plot of the 3D model shows 95.4% residues are present in the core region, 4.3% in allowed region, and 0.3% in disallowed region. Result of Verify_3D

shows that > 80% of the residues have an averaged 3D/1D score greater than 0.2 and ERRAT plot gives an overall quality factor of 98.428 to the modeled structure [10].

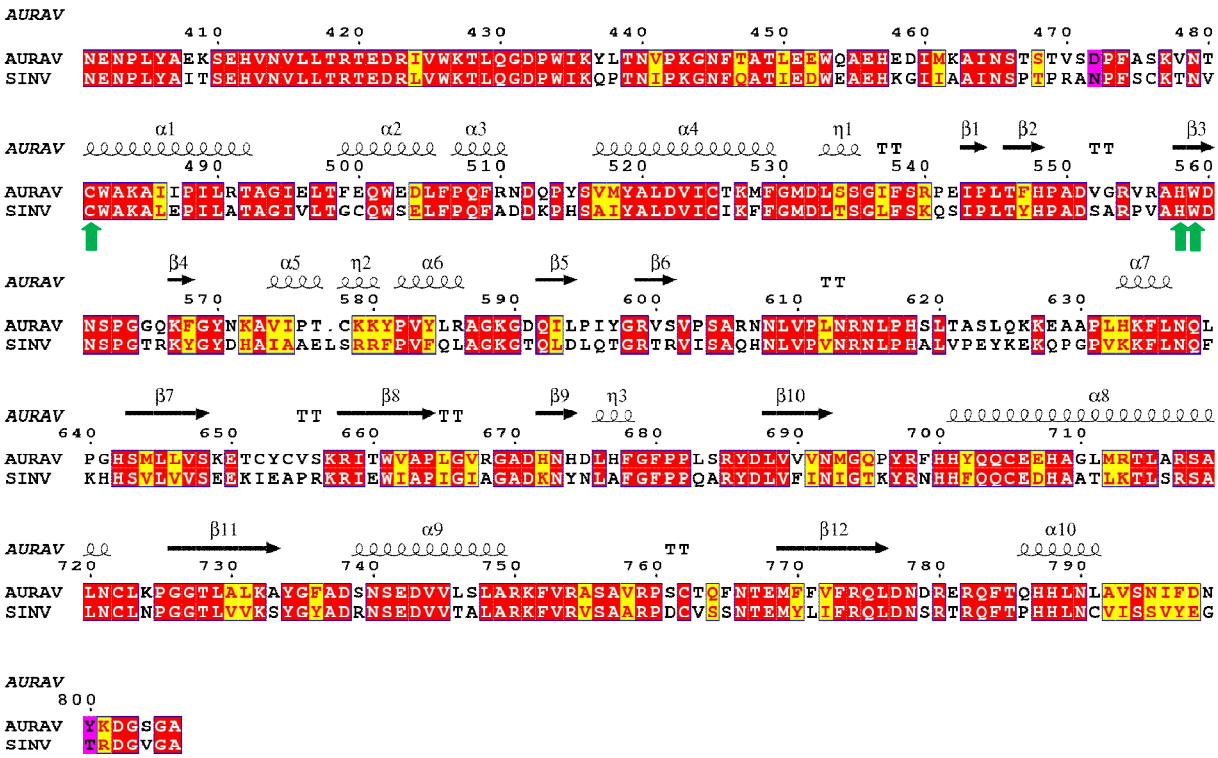
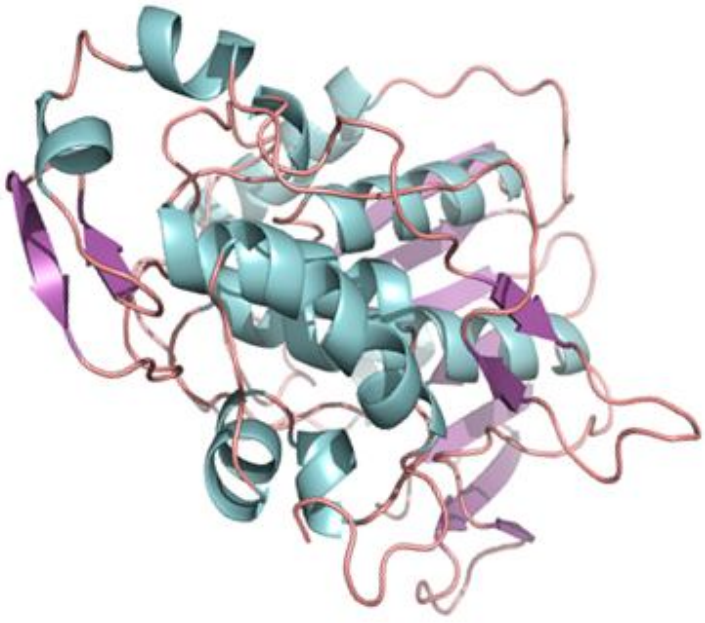


Figure 3.5 Sequence alignment of nsP2Aura protease with nsP2SIN. Identical amino acid residues are highlighted in red color background where as similar residues are highlighted in yellow color background. The secondary structural elements of nsP2AURA are shown above the aligned sequences. The conserved catalytic amino acid residues C481, H558, W559 have been indicated by green arrows. Made by ESPript[18].



A

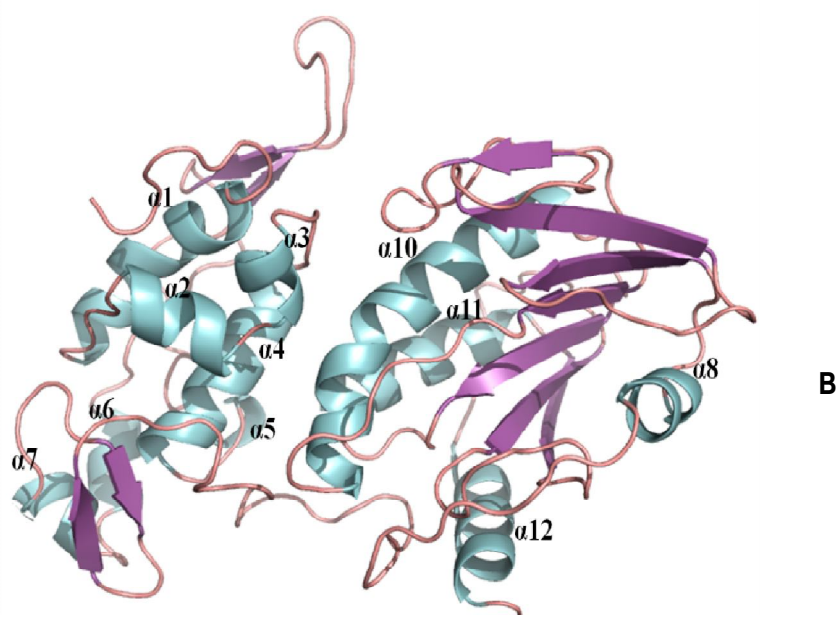


Figure 3.6 Cartoon diagram of nsP2AURA molecular model. (A) Top view of molecular model of nsP2AURA developed by Modeller *9v13*. **(B)** Side on view of the nsP2AURA. Helices are shown by cylindrical blue spiral ribbons and β-strands by purple arrows. Figure made by PyMoL [22].

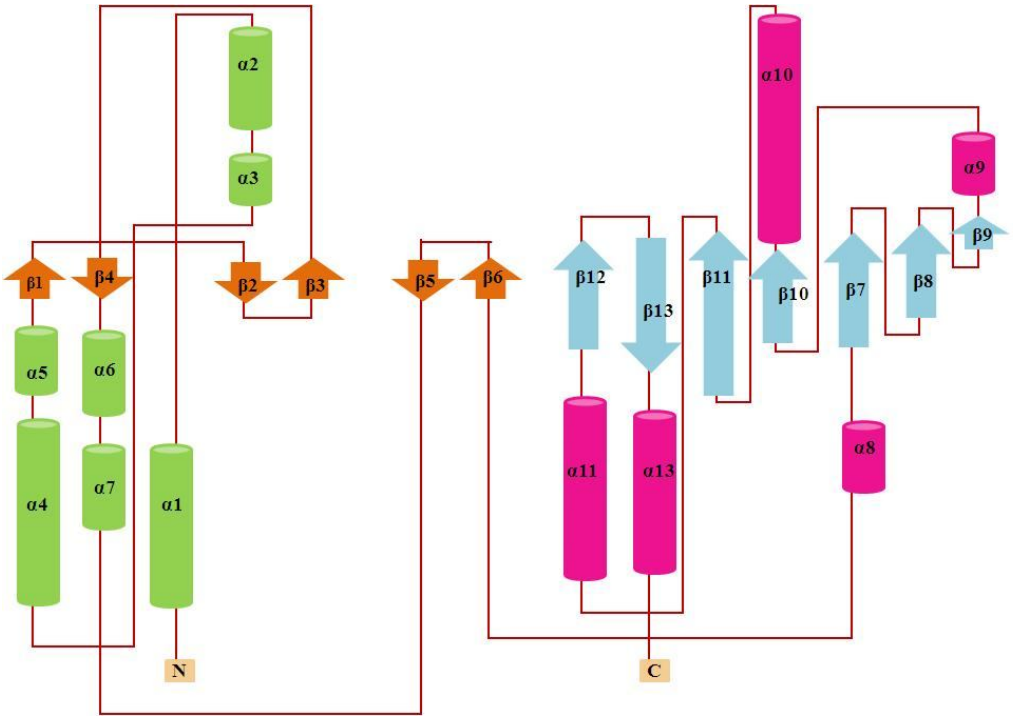


Figure 3.7 Topology representation of nsP2Aura monomer. The diagram illustrates how the two domains, N and C terminal connected by loops. The diagram also shows the relative

locations of the α-helices, here represented by cylinders and β-sheets represented by arrows. The small arrow indicates the directionality of the protein chain, from the N- to the C-terminus.

Inspite of only 18% similarity in the C terminal protease region, the active site residues, Cys481 and His558, W559 are conserved among all of the alphavirus nsP2 sequences. Interestingly the D560, N561, P562 residues (residue number in case of AURAV) following these actives site residues are also found to be conserved in all alphavirus (see Figure 5.10). And also the secondary structure elements are conserved even though there is limited sequence similarity in these regions. When the structure is analyzed, it gets clear that nsP2AURA belongs to the cysteine protease CA peptidase clan; with a papain-like fold. The polypeptide chain of the nsP2AURA folds into two distinct compact domains of approximately equal size with interface consisting regions made of helix and random coil.

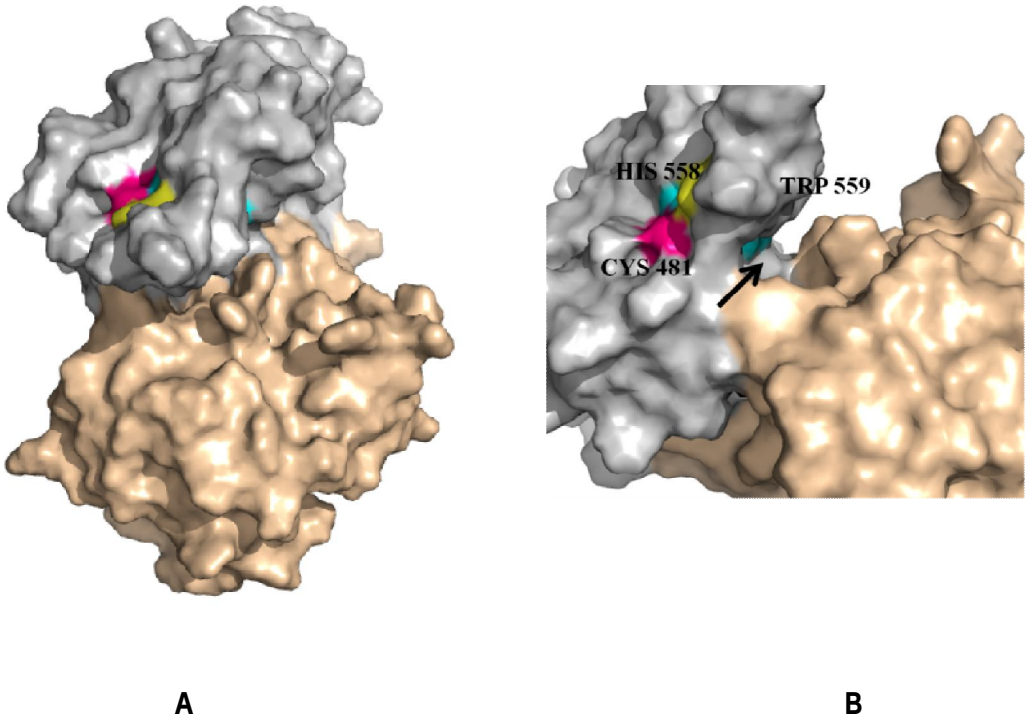


Figure 3.8 Active site of nsP2Aura. **(A)** Showing the residues Cys481 (Magenta), His558 (Yellow) and Trp559 (Cyan). **(B)** Substrate binding groove is marked by the arrow. The active site is at the juncture of the N and C terminal domain of the protein.

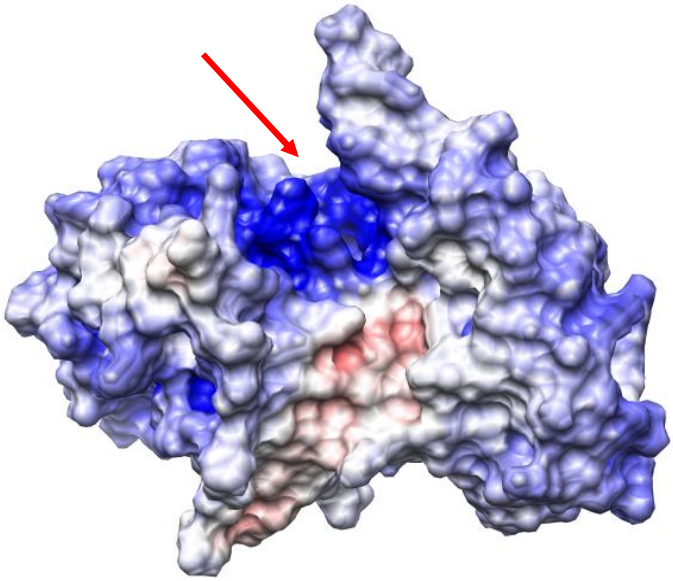


Figure 3.9 Electrostatic surface representation of the nsP2Aura. The active site cleft indicated by an arrow. The protein surface is colored according to its electrostatic potential from regions of red (negative potential) to blue (positive potential).

The N-terminal proteolytic domain from residue Asp472 to residue Asn613 contains the conserved protease catalytic dyad formed by the active site residues Cys481 and His558. It is organized around a central cluster of helices that are flanked by two short β hairpins. The orientation of the catalytic dyad residues in nsP2 is similar to the catalytic dyad conformation observed in SINV (Figure 3.10 B).

The catalytic cysteine is positioned at the N-terminal end of an α-helix, and the catalytic histidine is part of a β-strand. Third residue playing possible role in the catalytic activity was established to be Trp559 aligned on the β3 strand with the catalytic His558. This Trp is important as it helps in the recognition of the P2 Gly in the recognition sites at which the nsP2 protease cleaves [26]. The active site of the nsP2Aura seems like a groove which is open and tunnel shaped. Another important feature of the cysteine proteases is the presence of oxyanion hole which stabilizes a deprotonated oxygen or alkoxide, often by placing it close to positively charged residues; directly stabilizing the transition state of a reaction via stabilizing the transition intermediate. The oxyanion hole in case of nsP2Aura is likely to be formed close to

the active site by the backbone amides from Thr480 and Cys481. Positioning of the hole is such that it is in close proximity to the carbonyl group of the substrate's scissile bond.

The C terminal domain found to be similar to the SAM (S-adenosyl-L-methoionine) dependant methyltransferase having a typical Rossman symmetry with 7 strands of the β sheet arranged in the β3- β2- β1- β4- β5- β6- β7 order sandwiched between helix layers (Figure 3.6 B). The C-terminal methyltransferase domain is described from residue Asn614 to Trp800.

SAM dependant methyltarnsferases belongs to class I of the family of methyltransferases. The Rossman fold in these methyltransferases play role in the binding with the SAM molecules. They are highly versatile enzymes which help in the transfer of methyl group to proteins, DNA and various other biomolecules [25]. However little sequence identity of the domain with the methyltansferases and also the residues aligning the SAM binding pocket conferring little similarity, have led to the proposal deeming it as non-functional [24]. However, this methyltransferase fold could be used as a scaffold to bind RNA elements that may regulate RNA synthesis and virus replication [23, 24].

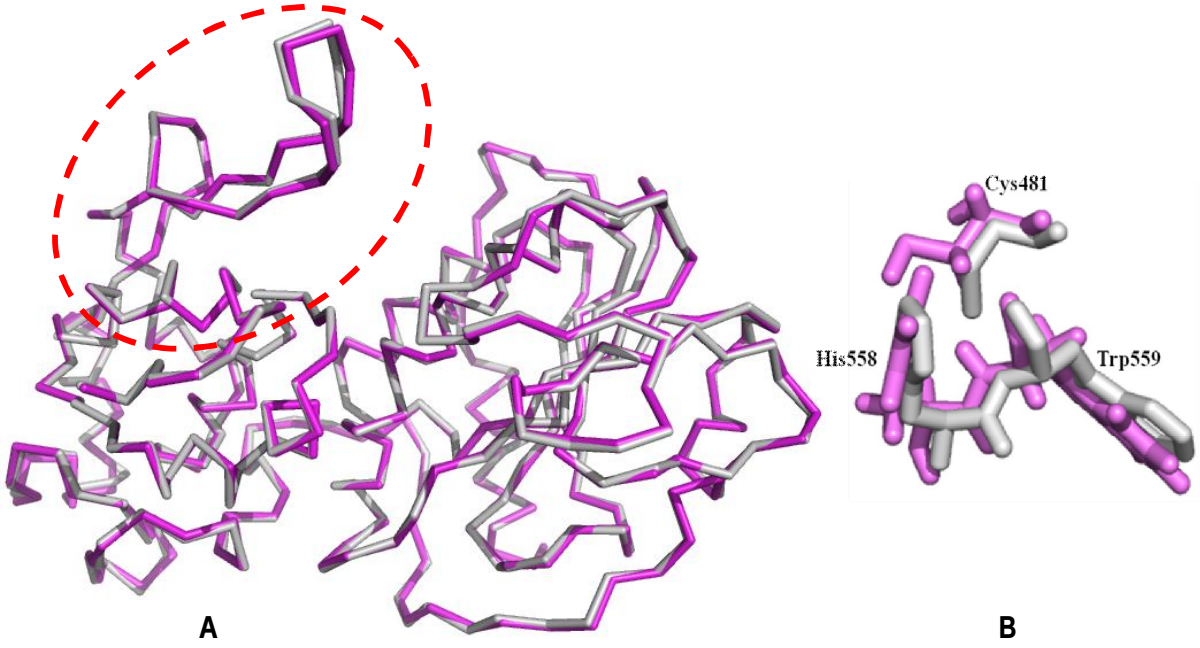


Figure 3.10 Superimposition of model and template structure. (A) Superimposition of **nsP2AURA** (pink) with **nsP2SIN** (grey) is shown in ribbon and active site region shown by dotted red circle. **(B)** Superimposition of active site residues of **nsP2AURA** (pink) and **nsP2SIN** (grey). Residue number for nsP2SIN C481, His558 and Trp559. Figures were made by PyMoL [22].

Comparing the nsP2AURA with its template nsP2SINV shows that they perfectly align over each other and a significant difference in main-chain conformation is present in residues Ala550-Ala547. The residues in this region are majorly conserved and variation in their main chain conformation indicates that this region is flexible and this flexibility could be of significance for the structure and activity of the enzyme. Both target and template proteins have highly-conserved cysteine and histidine residues with Cys481-His558 residues of nsP2Aura (pink in Fig.3.10 A) located approximately in the same spatial position as the Cys478-His556 residues in the active site of SINnsP2 (grey in Fig. 3.10 A). The distance between the Cys481 and His558 catalytic residues of SINnsP2 was found to be 3.3 Å (Figure 3.10 B), and also, the distance of 3.2 Å between the corresponding Cys481-His558 active site residues in the nsP2AURA model are comparable with the active site of the SINnsP2 template. It is therefore plausible that the nsP2AURA Cys481-His558 highly conserved residues have a similar catalytic mode of action.

3.5.4 β-galactosidase based activity assay

3.5.1 Cloning and expression of β-galactosidase gene with the nsP1/2, nsP2/3 and nsP3/4 sites

For developing β-galactosidase as the substrate for nsP2AURA, modifications were to be done for inserting the cleavage recognition sites into the β-galactosidase gene. The sites where nsP2AURA cleaves YDDAGAALVETP (nsP1/2 site), KDGSGAAPSYRV (nsP2/3 site), and QYLTGVGGYIFS (nsP3/4 site) were added into the β-galactosidase gene with the help of oligonucleotides. On the basis of information available in the literature, site on β-galactosidase gene was selected where this addition would be done so that the activity of the gene remains unaffected [19, 20]. So two step amplification of the β-galactosidase gene was done to incorporate nsP2 cleavage site (Figure 3.11), first set of primers amplified the gene from amino acid 1-578 from pSV template plasmid. The primers added *Nde*I and *Bam*HI sites to the ends of this segment. The amplified segment was then double digested with *Nde*I and *Bam*HI and gel extracted using Qiagen gel extraction kit. Also, pET-28c plasmid was double digested and gel extracted using the Qiagen gel extraction kit. The β-galactosidase segment I and pET-28c was then incubated with T4 DNA ligase for 15 hrs at 16°C. The ligated products were transformed in

DH5a cells using CaCl2 heat shock method. The transformed colonies grown on LB agar (with 50 µg/mL kanamycin) media were screened for the cloned β-galactosidase gene. The colonies were grown in 5 mL media LB broth with 50 µg/mL kanamycin and plasmid was isolated using Qiagen miniprep kit. The clone was confirmed by the PCR amplification using the isolated plasmids as the templates (Figure 3.12 A). The isolated plasmids were then digested with *Nde*I and *Bam*HI restriction enzymes and desired gene fragment was seen near 1600 bp on 0.8% agarose gel (Figure 3.12 B). Once the cloning of first segment of the β-galactosidase was confirmed, second segment of the β-galactosidase was amplified using the pSV plasmid as template. Three forward primers were used which were used in combination with reverse primer. Firstly, amplification was done using forward primers P1 and reverse primer (Table 3.1), the amplified product was used as a template for amplification with forward primer P2 and reverse primer. And the amplified product of this reaction was used as template for amplification with forward primer P3 and reverse primer. This final product is assumed to be having the nsP2Aura cleavage site at the 5' end of the segment added with the help of three different forward primers. The amplified gene was also double digested with *Bam*HI and *Xho*I sites added on its ends with the help of the primers.

The digested segment was then gel extracted using Qiagen gel extraction kit. Also the pET-28c plasmid containing β-galactosidase segment (1-578) previously cloned was double digested with *Bam*HI and *Xho*I enzymes and gel extracted. Both β-galactosidase segment II and pET-28c- β-galactosidase-SegmentI (1-578) were then incubated at 16 °C for 15 hrs in the presence of T4 DNA ligase. The ligated products were transformed in DH5a cells using CaCl2 heat shock method. The transformed colonies grown on LB agar and kanamycin media were screened for the cloned β-galactosidase gene. The colonies were grown in 5 mL media LB broth with kanamycin and plasmid was isolated using Qiagen miniprep kit. The clone was confirmed by the PCR amplification using the isolated plasmids as the templates (Figure 3.13 A-C). The isolated plasmids were then digested with *Bam*HI and *Xho*I restriction enzymes and desired gene fragment was seen near 1400 bp on 0.8% agarose gel (Figure 3.13 D). The selected constructs were sequenced in both the directions using vector specific T7 promoter universal primers by dye termination method by IDT Biosolutions Company.

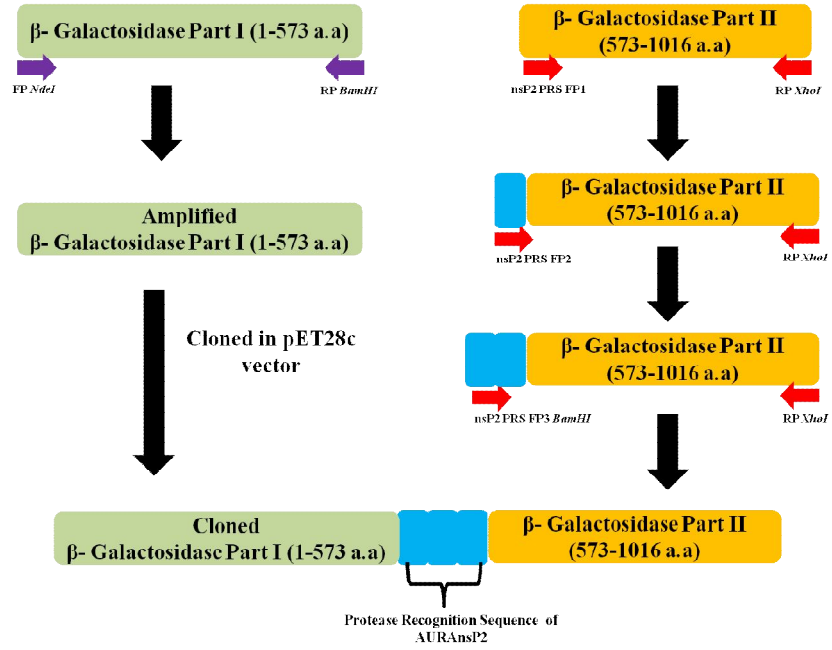


Figure 3.11 Cloning strategy of β-galactosidase: Firstly cloning of first part of β-galactosidase was done in pET-28c vectors by using sequence specific primers. Second step involves cloning of second part of B-galactosidase having PRS of nsP2AURA.

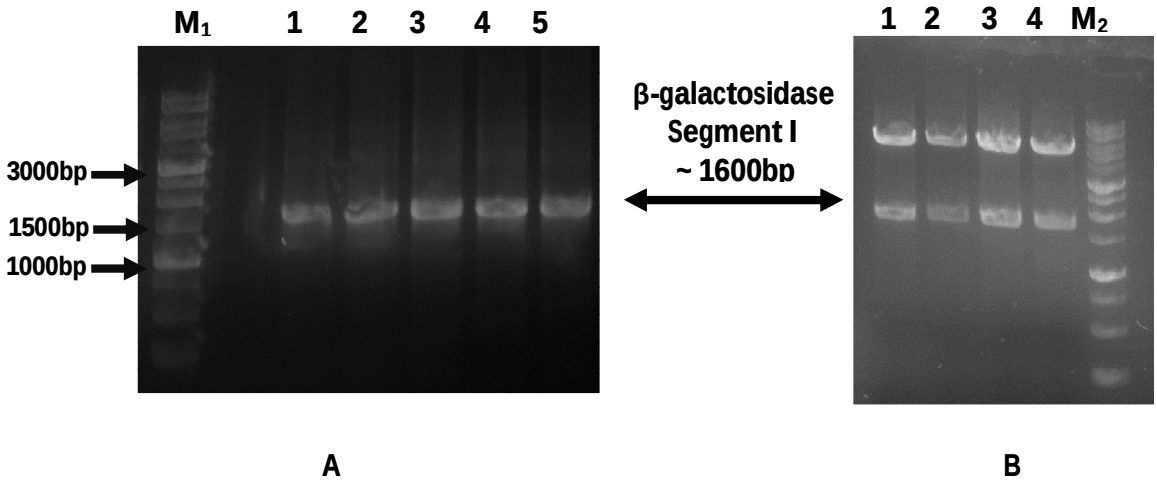


Figure 3.12 Amplification and cloning of β-galactosidase-nsP2AURA site gene. Agarose gel electrophoresis (A) PCR products. lane M1: DNA marker; lane 1-5: PCR product amplified from pSV plasmid with gene specific primers. **(B)** Recombinant plasmids digested with *NdeI*

and *BamHI*. lane M2: DNA marker; lane 1: pET-28c-β-galactosidase segment I from the screened colonies.

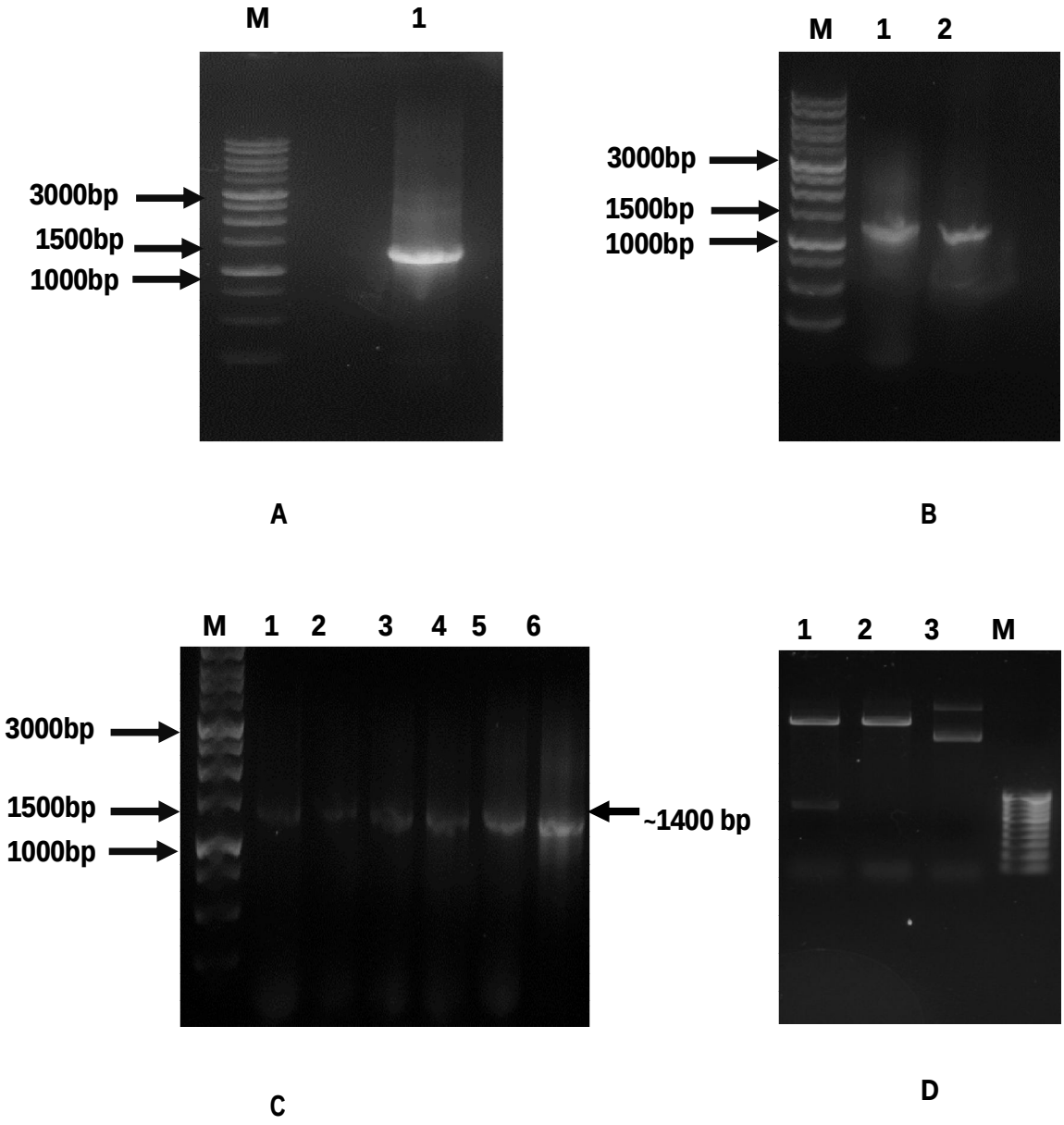
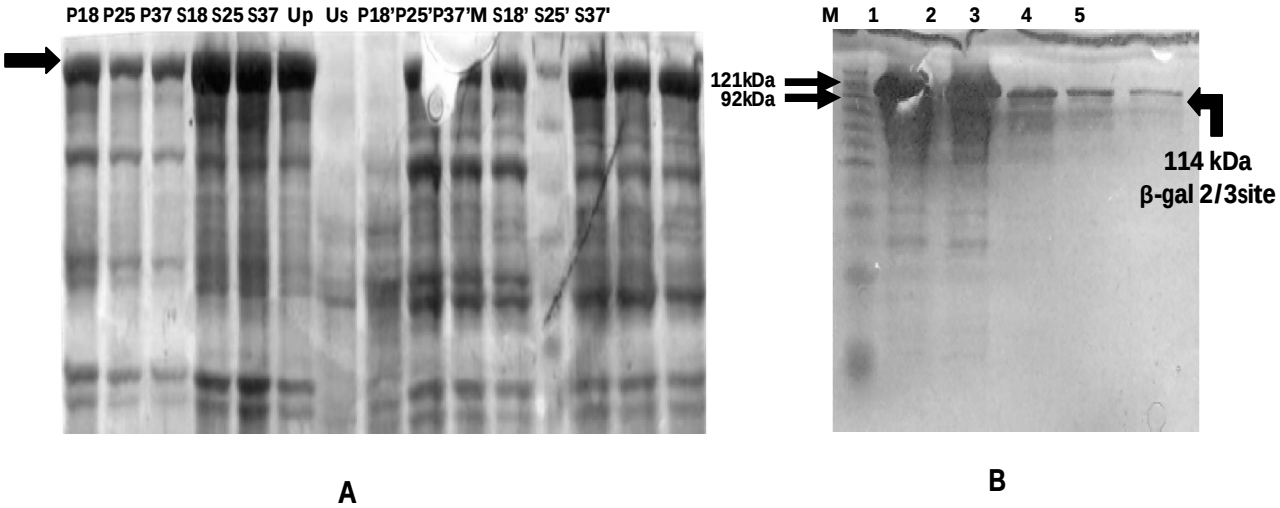


Figure 3.13 Amplification and cloning of β-galactosidase-nsP2AURAsite gene with nsP2AURA 1/2, 2/3 and 3/4 recognition sites. Agarose gel electrophoresis **(A-C)** PCR products amplified from pSV plasmid with gene specific primers. Using P1, P2 and P3 forward primers **(D)** Recombinant plasmids digested with *Bam*HI and *Xho*I. Lane M: DNA marker; lane 1-3: pET-28c-β-galactosidase segment II from the screened colonies.

For expression of the recombinant β-galactosidase, the cloned pET-28c vector was transformed in to *E. coli BL-21* (DE3) cells. The small scale expression of the recombinant protein was analyzed at three different temperatures (18°°C, 25°C and 37°°C). Optimum expression and solubility of recombinant β-galactosidase was observed at all the temperatures. So, 37 °C temperature and induction with 0.4 mM IPTG concentration was selected as the ideal condition for the expression of the recombinant protein.

3.5.2 Purification of β-galactosidase^{R} by affinity chromatography

For purification of pET-28c- β-galactosidase^{R} one liter culture of BL-21 (DE3) cells, containing pET-28c-β-galactosidase^{R} were grown for 4 hrs at 37 °C. Each liter of cells was grown in a 2 litre baffled flask at 200 rpm. Following growth cells were pellet down by centrifugation at (3412g, 10 min.) and resuspended in 30 mL of buffer A (50 mM phosphate pH 7.5, 5% Glycerol, 500 mM NaCl) containing 0.5 mg/mL of lysozyme and 0.01 mg/mL of DNaseI. The cells were incubated for 10 min. and then lysed using French press (Constant cell Disruptor Systems, UK). The disrupted cells were then subjected to centrifugation so as to separate the soluble portion from the insoluble fraction (at 28928g, 60 min., 4°C). Since the protein was having 6xhis tag at the N-terminal end, the purification was easily achieved by the His trap HP column, having nickel which has affinity towards 6xHis.



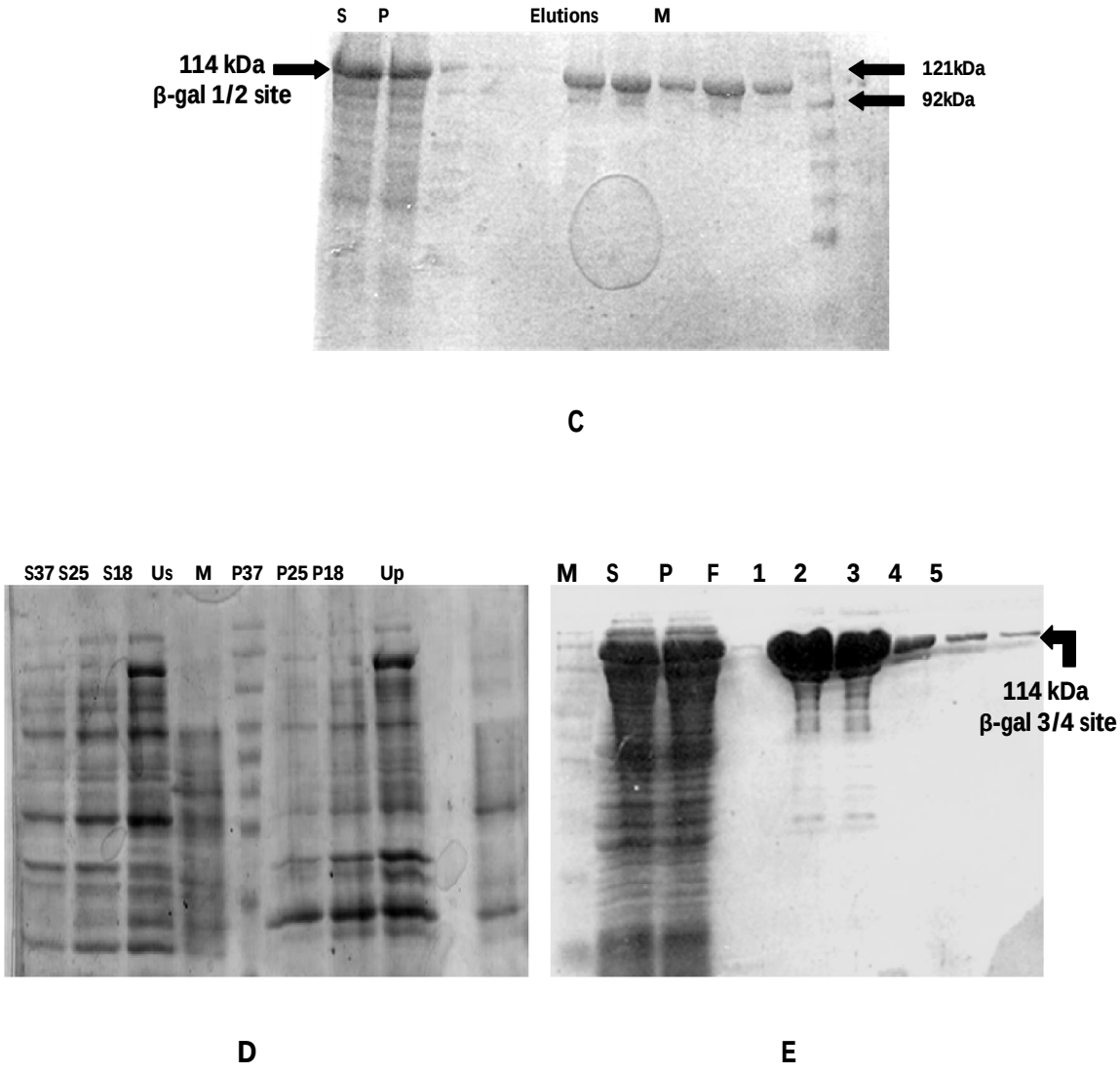


Figure 3.14 Expression analysis and purification of β-galactosidase; **(A)** Expression of β-galactosidase 1/2 and 2/3 at different temperatures; S18, S25, S37, P18, P25 and P37 supernatant and pellet of β-gal with nsP2AURA 1/2 cleavage site and S18', S25', S37', P18', P25' and P37' supernatant and pellet of β-galactosidase with nsP2AURA 2/3 cleavage site along with uninduced supernatant (Us) and pellet (Up). **(B)** Purification of β-galactosidase with nsP2 1/2 site using affinity chromatography. **(C)** Purification of β-galactosidase with nsP2 2/3 site using affinity chromatography. (**D)** Expression of β-galactosidase with 3/4 site at different temperatures; S18, S25, S37, P18, P25 and P37 supernatant and pellet of β-gal with nsP2AURA 3/4 cleavage site. **(E)** Purification of β-galactosidase with nsP2 3/4 site using affinity chromatography.

Thus the His trap HP column was equilibrated with the buffer A. The cleared cell lysate was then loaded on the equilibrated column at the flow rate of 0.4 mL/min. Protein was eluted with 10x column volume linear gradient from 0% to 100% buffer B (50 mM phosphate pH 7.5, 5% glycerol, 250 mM NaCl), at the rate of 2 mL/min followed by isocratic wash with 100% buffer B. 5 mL fractions were collected over the course of gradient and isocratic wash. Fractions containing the pET-28c-β-galactosidase^{R} were pooled (based on the absorbance peak) and dialysed against the dialysis buffer I (50 mM phosphate pH 7.5, 5% Glycerol, 0.5 mM DTT). Sample from fractions were run on 12% SDS-PAGE gel to analyze the protein purity (Figure 3.14 B, C, E). The dialysed protein was then concentrated using Amicon (10kDa cut-off), only after confirmation by loading on the SDS-PAGE. Protein was concentrated to around 20 mg//mL. This strategy of purification was used for the purification of all the three β -galactosidase^{R} 1/2, 2/3 and ¾.

3.5.3. β -galactosidase based activity assay

β-galactosidase is encoded by the lacZ gene of the lac operon in E. coli. It is a large (120 kDa, 1024 amino acids) protein that forms a tetramer. The enzyme's function in the cell is to cleave lactose to glucose and galactose so that they can be used as carbon/energy sources. The synthetic compound o-nitrophenyl-β-D-galactoside (ONPG) is also recognized as a substrate and cleaved to yield galactose and o-nitrophenol which has a yellow color (Figure 3.15). When ONPG is in excess over the enzyme in a reaction, the production of o-nitrophenol per unit time is proportional to the concentration of β-galactosidase; thus, the production of yellow color can be used to determine enzyme concentration.

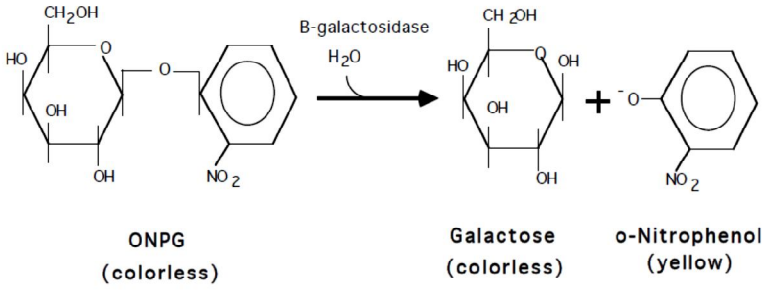


Figure 3.15 Mechanism of β-galactosidase activity.

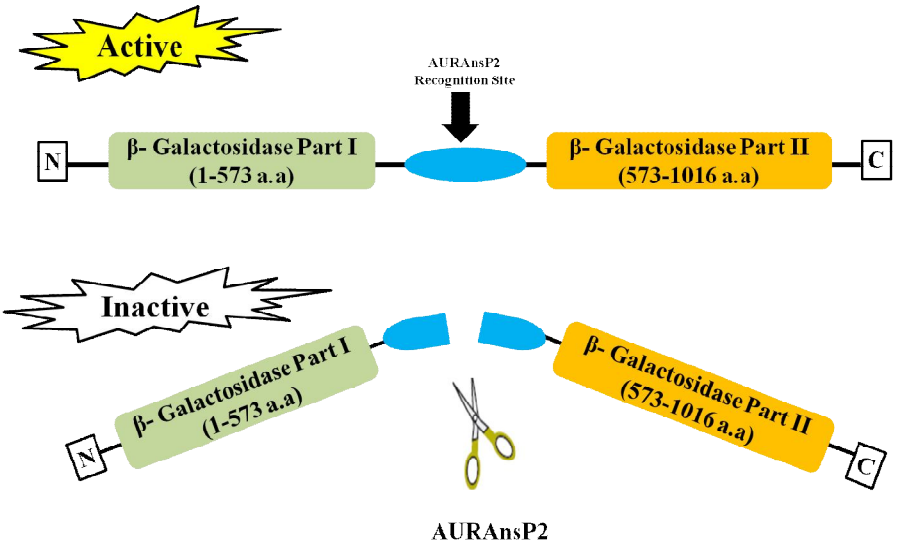


Figure 3.16 nsP2 activity assay using β-galactosidase^{R}. Cleavage at the PRS site makes the β-galactosidase inactive and in the presence of ONPG no reaction takes place.

Before optimizing the assay for high throughput screening, initial testing whether the activity change is due to the cleavage at the PRS (Protease (nsP2) recognition sequence) or not, the assay was performed in disposable plastic cuvettes where the reaction was set in the different ratios of nsP2AURA with the substrate β-galactosidase. The purified pET-28c- β-galactosidase^{R} was incubated with the purified nsP2AURA in 2:1 molar concentrations at different temperature conditions i.e from 20°C to 45°C. The incubated mixtures were mixed by 6x loading buffer and loaded on 10% SDS-PAGE gel.

The gel was stained and destained using the standard procedure. In temperatures above the 37°C mostly aggregation was observed and at lower temperatures the rate of cleavage was slow. So finally 37 °C was selected as the optimum temperature for the reaction. The reaction was done by incubating the nsP2AURA protease with only the β-galactosidase 3/4 since in case of other modified β-galactosidases^{R} (2/3 and 1/2) there was negligible cleavage (Figure 3.17). After cleavage two different fragments of size ~62 and 48 kDa are observed (Figure 3.17 C).

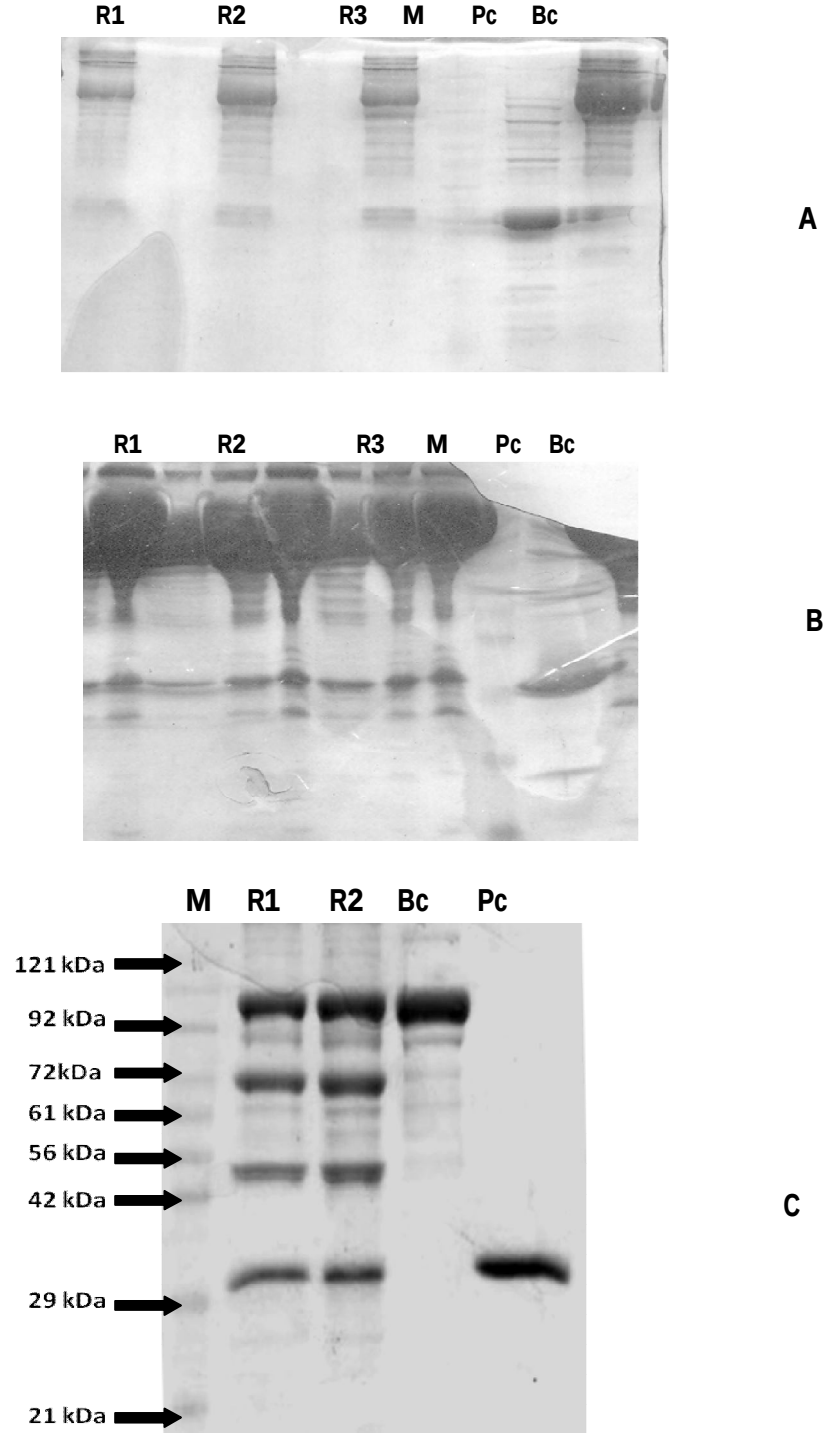


Figure 3.17 Cleavage of β-galactosidase^{R} 1/2, 2/3, 3/4 site with nsP2AURA protease; analysis on SDS-PAGE. M is the Protein Ladder, R1 – Reaction Mix 1, R2 – Reaction Mix 2 (consisting of β-galactosidase and nsP2AURA protease), βc – β-galactosidase^{R} control, Pc – nsP2AURA Protease control. **(A)** Cleavage of β-galactosidase^{R} containing 1/2 site. **(B)** Cleavage of β-galactosidase^{R} containing 2/3 site. **(C)** Cleavage of β-galactosidase^{R} containing 3/4 site.

The experiment was done in triplicate by repeating the above conditions of the reaction in the microtitre plate. The assay was performed by incubating the pET-28c-β-galactosidase^{R} (containing nsP2AURA 3/4 site) in the Z buffer at different molar concentrations (Figure 3.16). After that added ONPG with final concentration of 10 µM was added to the well and the mix was incubated at 37 °C for 30 min. The reaction was halt by the addition of 100 µL of 1M calcium carbonate which changes the pH of the reaction so as to stop it. The readings were taken at 420 nm at which the O-nitrophenol absorbs. Then the experiment was performed by fixing the substrate i.e pET-28c-β-galactosidase^{R} concentration with varying concentration of nsP2AURA.

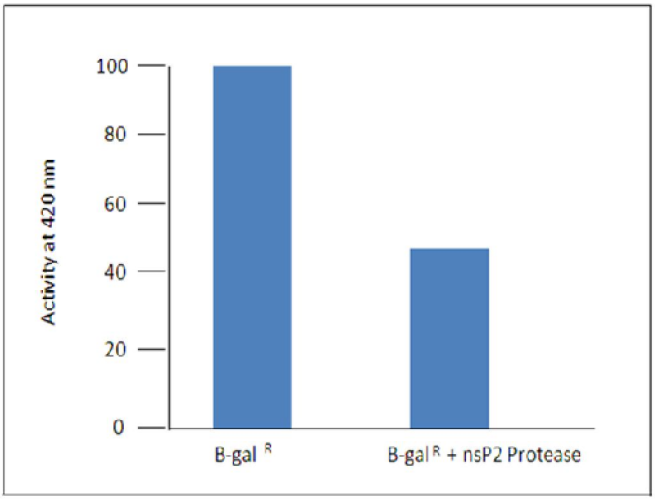


Figure.3.18 Change in activity of β-galactosidase ^{R} 3/4 in the presence of nsP2AURA protease. nsP2 protease reduced the activity of the β-galactosidase ^{R} 3/4 upto 50 percent by cleaving at the nsP3/4 recognition site inserted inside the β-galactosidase gene.

3.6 Conclusion

nsP2AURA was cloned and expressed in a bacterial expression system. Over expressed protein was purified using ion exchange and size exclusion chromatography. However our initial thoughts were to crystallize the protein and understand the evolutionary features which the nsP2 protein might have undergone considering the fact that AURA virus is observed to be a connecting link between the Old and New World of Alphaviruses. But when those attempts failed, we deduced the structure of the protease domain using structure based modeling. We were able to conclude that there are regions which are different in nsP2AURA which are different from its template i.e SIN nsP2. However we cannot conclude exactly about the deviations which the protein might have undergone and what are the other residues (besides active site) which might be necessary for the important functions. So, studies can be done to establish the relation between the structure and the evolution which the virus might have undergone during the course of time.

Also, we designed a simple yet efficient assay for the screening of the inhibitors against the alphavirus nsP2 protease. The β-galactosidase based assay is economical and would help in the screening the inhibitors against the protease.

3.7 Bibliography

1. Lundström, Jan O., and Martin Pfeffer. "Phylogeographic structure and evolutionary history of Sindbis virus." *Vector-Borne and Zoonotic Diseases* 10, no. 9 (2010): 889-907.
2. Rümenapf, Tillmann, Ellen G. Strauss, and James H. Strauss. "Subgenomic mRNA of Aura alphavirus is packaged into virions." *Journal of virology* 68, no. 1 (1994): 56-62.
3. Rümenapf, Tillmann, Ellen G. Strauss, and James H. Strauss. "Aura virus is a New World representative of Sindbis-like viruses." *Virology* 208, no. 2 (1995): 621-633.
4. Calisher, Charles H., Nick Karabatsos, John S. Lazuick, Thomas P. Monath, and Katherine L. Wolff. "Reevaluation of the western equine encephalitis antigenic complex of alphaviruses (family Togaviridae) as determined by neutralization tests." *The American journal of tropical medicine and hygiene* 38, no. 2 (1988): 447-452.
5. Hahn, Chang S., Shlomo Lustig, Ellen G. Strauss, and James H. Strauss. "Western equine encephalitis virus is a recombinant virus." *Proceedings of the National Academy of Sciences* 85, no. 16 (1988): 5997-6001.
6. Weaver, Scott C., Wenli Kang, Yukio Shirako, Tillman Rumenapf, Ellen G. Strauss, and James H. Strauss. "Recombinational history and molecular evolution of western equine encephalomyelitis complex alphaviruses." *Journal of virology* 71, no. 1 (1997): 613-623.
7. Levinson, Randy S., James H. Strauss, and Ellen G. Strauss. "Complete sequence of the genomic RNA of O'nyong-nyong virus and its use in the construction of alphavirus phylogenetic trees." *Virology* 175, no. 1 (1990): 110-123.
8. Russo, Andrew T., Mark A. White, and Stanley J. Watowich. "The crystal structure of the Venezuelan equine encephalitis alphavirus nsP2 protease." *Structure* 14, no. 9 (2006): 1449-1458.
9. Webb, Benjamin, and Andrej Sali. "Protein structure modeling with MODELLER." In *Protein Structure Prediction*, pp. 1-15. Springer New York, 2014.
10. Eisenberg, David, Roland Lüthy, and James U. Bowie. "VERIFY3D: assessment of protein models with three-dimensional profiles." *Methods in enzymology* 277 (1997): 396-404.
11. Mandel, Morton, and Akiko Higa. "Calcium-dependent bacteriophage DNA infection." *Journal of molecular biology* 53, no. 1 (1970): 159-162.
12. Kruger, Nicholas J. "The Bradford method for protein quantitation." In *Basic protein and peptide protocols*, pp. 9-15. Humana Press, 1994.

13. Shirako, Yukio, Bo Niklasson, Joel M. Dalrymple, Ellen G. Strauss, and James H. Strauss. "Structure of the Ockelbo virus genome and its relationship to other Sindbis viruses." *Virology* 182, no. 2 (1991): 753-764.
14. Ou, Jing-Hsiung, Dennis W. Trent, and James H. Strauss. "The 3′-non-coding regions of alpharivus RNAs contain repeating sequences." *Journal of molecular biology* 156, no. 4 (1982): 719-730.
15. Weaver, Scott C., Liz Anne Bellew, and Rebeca Rico-Hesse. "Phylogenetic analysis of alphaviruses in the Venezuelan equine encephalitis complex and identification of the source of epizootic viruses." *Virology* 191, no. 1 (1992): 282-290.
16. Rentier-Delrue, Francoise, and Nathaniel A. Young. "Genomic divergence among Sindbis virus strains." *Virology* 106, no. 1 (1980): 59-70.
17. McWilliam, Hamish, Weizhong Li, Mahmut Uludag, Silvano Squizzato, Young Mi Park, Nicola Buso, Andrew Peter Cowley, and Rodrigo Lopez. "Analysis tool web services from the EMBL-EBI." *Nucleic acids research* 41, no. W1 (2013): W597-W600.
18. Robert, Xavier, and Patrice Gouet. "Deciphering key features in protein structures with the new ENDscript server." *Nucleic acids research* 42, no. W1 (2014): W320-W324.
19. Feliu, Jordi X., and Antonio Villaverde. "Engineering of solvent-exposed loops in Escherichia coli β-galactosidase." *FEBS letters* 434, no. 1 (1998): 23-27.
20. Juers, Douglas H., Brian W. Matthews, and Reuben E. Huber. "LacZ β-galactosidase: Structure and function of an enzyme of historical and molecular biological importance." *Protein Science* 21, no. 12 (2012): 1792-1807.
21. Shin, Gyehwa, Samantha A. Yost, Matthew T. Miller, Elizabeth J. Elrod, Arash Grakoui, and Joseph Marcotrigiano. "Structural and functional insights into alphavirus polyprotein processing and pathogenesis." *Proceedings of the National Academy of Sciences* 109, no. 41 (2012): 16534-16539.
22. DeLano, Warren L. "The PyMOL molecular graphics system." (2002).
23. Mayuri, Geders, Todd W., Janet L. Smith, and Richard J. Kuhn. "Role for conserved residues of Sindbis virus nonstructural protein 2 methyltransferase-like domain in regulation of minus-strand synthesis and development of cytopathic infection." *Journal of virology* 82, no. 15 (2008): 7284-7297.

24. Sawicki, Dorothea L., Silvia Perri, John M. Polo, and Stanley G. Sawicki. "Role for nsP2 proteins in the cessation of alphavirus minus-strand synthesis by host cells." *Journal of virology* 80, no. 1 (2006): 360-371.
25. Struck, Anna-Winona, Mark L. Thompson, Lu Shin Wong, and Jason Micklefield. "S-Adenosyl-Methionine-Dependent Methyltransferases: Highly Versatile Enzymes in Biocatalysis, Biosynthesis and Other Biotechnological Applications." *Chembiochem* 13, no. 18 (2012): 2642-2655.
26. Golubtsov, Andrey, Leevi Kääriäinen, and Javier Caldentey. "Characterization of the cysteine protease domain of Semliki Forest virus replicase protein nsP2 by in vitro mutagenesis." *FEBS letters* 580, no. 5 (2006): 1502-1508.
27. Gish, Warren, and David J. States. "Identification of protein coding regions by database similarity search." *Nature genetics* 3, no. 3 (1993): 266-272.

Chapter IV:

Sindbis nsP2 papain-like protease expression, purification, characterization and crystallization

4.1 Introduction

Belonging to *togaviridae* family of virus classification, Sindbis virus is the model virus and representative of the Old world group of Alphaviruses [2, 3]. Sindbis virus is positive sense RNA virus of around 11.7 kb genome, belonging to western equine encephalitis complex of the *Alphavirus* genus [4, 5]. The virus was first discovered in the year 1952 in the Nile river delta where it was isolated from Culex (*Culex pipiens and Cx. univittatus*) mosquitoes [6]. Subsequently, first human case was reported in the Uganda in 1961 [7], followed by South Africa in 1963 [8] and Australia in 1967 [9]. The virus has been identified in the fauna in the following countries: Czech Republic, Finland, Germany, Norway, Romania, Serbia, Spain, Sweden, Australia, Ukraine, and in the UK [11]. In countries like Finland which is most affected by Sindbis, sporadic cases are reported every year and after every seven year larger outbreaks occur targeting masses [1, 13]. Despite the large geographical distribution of Sindbis, cases of human infection have been mainly reported from northern European and some South African countries [12].

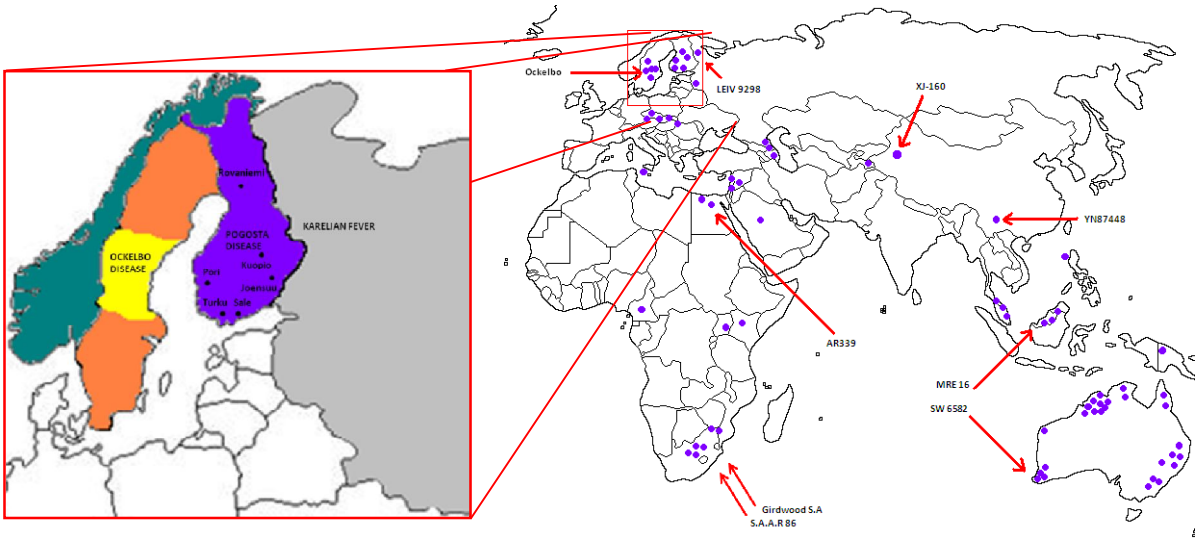


Figure 4.1 Epidemology of SINV and different strains isolated from different vertebrates and/or mosquito species from the Old world. In box endemic areas of SINV infection. Based on data from ref. [1].

Culex and Culiseta mosquitoes are the carrier vectors of the Sindbis virus which help in the spread and the maintenance of the virus in the environment. SINV circulates between ornithophilic Culex mosquito species and birds. Birds are the important reservoirs for SINV, which even play role in transmission of the virus to the longer areas [15, 16,]. Humans, mammals and different bird species have been reported to be seropositive of Sindbis virus [11, 17, 18, 19]. Sindbis is acknowledged to be the causative agent of a rash-arthritic syndrome [20]. The disease caused by the virus is known as Pogosta in Finland, Ockelbo disease in Sweden, and as Karelian fever in Russia [21, 22, 23]. In SINV infection, initial incubation period lasts for less than seven days with symptoms like mild fever, pain in joints particularly in hips, knees, wrists and ankle, sometimes accompanied by maculopapular and itchy exanthema [21, 24, 25] over the limb and trunk region. The disease is usually mild in case of children and rarely leads to the arthritis like symptoms but in adults is observed the severe form of the infection which would persist for few months to years and develop into chronic arthritis [25, 26]. Lately no specific medications or drugs are available against the SINV infection and only preventive measures along with symptomatic therapy come in handy. Prevention includes controlling the brooding population of mosquitoes in the affected areas and covering the maximum body parts along with application of mosquito repellant in exposed areas of the body. Treatment includes usage of antihistamins for rashes, and intra-articular cortocosteriods and non-salicylate analgesics for persistent join pain [1, 28]. Detection kits for SINV infection are based on ELISA platform using detection of anti-SINV inmmunoglobulins M (IgM) or inmmunoglobulins (IgG) antibodies, detectable with 7-10 days of infection [27]. Immunofluorescence and Hemagglutination [1] based inhibition techniques are also used for detection of anti-SINV antibodies.

With the information about SINV infection coming handy in case of emergency, there is still no effective medication or vaccine available in case an infection of epidemic scale occurs. So, we tried to target the non structural protease from this virus also. Although SINV does belongs to the Old World of Alphaviruses like CHIKV and AURA but it has some distinctions from these members of the family. Our idea of targeting the protease was to extract the structural information of the protease from SINV so as to understand the mechanism of its functioning in this Old world member.

4.2 Cloning and expression of pET-28c-nsP2SINV (472-801) (with 6xhis tag), pET-28c-nsP2SINV (472-603), pET-28c-nsP2SINV (472-609)

4.2.1 Material

4.2.1.1 Viral sequences and Plasmids

The coding sequences for nsP2 from SINV were amplified by PCR from cDNA [pToto64] plasmid obtained from Prof. Richard Kuhn's laboratory. *Escherichia coli DH5α* and *Rosetta* (DE3) strains and cloning and expression plasmids were purchased from Novagen (USA).

4.2.1.2 Chemicals

All chemicals and reagents used in this study were of ACS grade and purchased either from Merck-millipore, Sigma, Fluka, and BioRad. Chemicals procured from other then these companies are listed below. Growth media were purchased from Himedia Laboratories, Mumbai, India. Chromatography media and columns were purchased from GE Healthcare.

The oligonucleotides for the PCR amplification of nsP2SINV gene were designed with the help of Oligoanalyser server of IDT technologies and purchased from Ocimum technologies.

4.2.1.3 Construct identification

Initial constructs were selected manually to replicate and optimize results reported for the sequences available at the time (SINV, SFV, RRV, CHIKV and VEEV). More extensive sequence analysis for construct selection utilized multiple sequence alignments of all available alphavirus sequences in public databases by the program CLUSTALW [28] in conjunction with the disorder prediction algorithm of FoldIndex Disorder prediction using the FoldIndex algorithm proved extremely powerful, and incorporation of that algorithm into the Imatrix software package extended FoldIndex [30] to use multiple sequence alignments instead of single sequences for disorder prediction. Using sequence conservation and disorder predictions as guides, various N- and C-termini of putative domains were tested to identify soluble constructs for defining domain boundaries.

4.2.1.4 Crystallization solutions

Crystallization screens (Crystal Screen I & II, PEG/ion I & II, Index, Salt and Crystal Screen Cryo) were procured from Hampton Research. For optimization reagents were made by highest purity ACS grade chemicals. To remove insoluble particles, solutions were filtered through 0.22 μ filter. The prepared reagents were maintained 4°C. PEG solutions were prepared by overnight stirring and stored in light protected bottles.

4.2.2 Methods

4.2.2.1 Cloning of nsP2SINV in pET expression vectors

Cloning and expression analysis of nsP2SINV 472-801 protease in pET-28c vector between *Nde*I and *Xho*I (6xhis tag) was done as described in the case of pET-28c-nsP2CHIKV case. Two different constructs nsP2SINV 472-603, and nsP2SINV 472-609 were developed for the protease domain excluding the C-terminal methyltransferase domain.

4.2.2.2 Optimization of expression conditions of nsP2SINV constructs

Plasmids encoding various constructs were transformed into *E. coli Rossetta* (DE3) with pRARE2 plasmid. The pRARE2 plasmid supplies tRNAs for 7 codons rarely used in bacterial genes (AGA, AGG, AUA, CUA, GGA, CCC, and CGG) driven by their native promoters. Colonies were used to inoculate 5 mL of LB liquid culture in 15 mL glass culture tubes and grown at 37°C to an O.D.600 of 0.4-0.5. The cultures were adjusted to 18 ° C, room temperature (~25°C) and 37°C. After thermal equilibration, IPTG was added to a final concentration of 0.4 mM and grown for 12-16 hrs, 6-8 hrs, or 3-4 hrs depending on induction temperature. Cultures were then harvested by centrifugation in a clinical centrifuge (3214 g for 10 min., room temperature) and pellets were resuspended in 500 μL of buffer A (50 mM Tris pH 7.5, 500 mM NaCl, 20 mM imidazole). Cells were lysed by sonication on ice (20 sec with microfuge tip) and spun at 28,928 g for 10 min. to separate insoluble material. 20 μL of soluble extract were added to 20 μL of 2X reducing SDS-PAGE sample buffer (2X = 50 mM Tris pH 6.8, 2% SDS, 0.1% bromophenol blue, 10% glycerol, 100 mM DTT) and boiled for 5 min. at 95°C. The pellet fraction was resuspended in 500 μL of 2X reducing SDS-PAGE sample buffer and boiled for 5 min.at 95°C. These gel samples (10 μL of soluble fraction gel sample, 5 μL of insoluble fraction gel sample) were then run on a 12% SDS-PAGE gel and stained with Coomassie stain.

4.2.2.3 Large scale expression and purification of pET-28c- nsP2SINV (472-801) (with 6xhis-tag).

Two liter culture of *rossetta* (DE3) cells, containing pET-28c- nsP2SINV (with N-terminal 6xhis-tagged) was grown for 20 hrs at 18°C. Each liter of cells was grown in a 2 liter baffled flask at 200 rpm. In general, an overnight culture was used to inoculate 1 litre of LB media in a 2 litre baffle flask and grown to an O.D.600 of 0.4-0.5 at 37°C. The temperature was then adjusted to 18°C and IPTG was added to a final concentration of 0.4 mM and allowed to grow 16-18 hrs with shaking. It should be noted that SINV nsP2 protease domain [472-801] was extremely sensitive to induction temperature; expression of SINV nsP2 protease domain above 20°C resulted in completely insoluble protein. Cells were harvested by centrifugation into 50 mL corning "Falcon" tubes and either processed immediately or frozen at -20°C. Following this the pellets were resuspended in 30 mL of buffer A (50 mM Tris pH 7.5, 5% glycerol, 500 mM NaCl) containing 0.5 mg/mL of lysozyme and 0.01 mg/mL of DNaseI (10 mM MgCl2). The cells were incubated for 10 min. and then lysed using French press (Constant cell Disruptor Systems, UK). The disrupted cells were then subjected to centrifugation so as to separate the soluble portion from the insoluble fraction (at 28,928 g, 60 min., 4°C). In the meantime the His trap HP column was equilibrated with the buffer A. The cleared cell lysate was then loaded on the equilibrated 5 mL HiTrap Chelating HP Ni-IDA resin (GE Healthcare) or 5 mL His Trap Ni-NTA resin (GE Healthcare) column at the flow rate of 0.4 mL/min. Protein was eluted with 10x column volume linear gradient from 0% to 100% buffer B (50 mM Tris pH 7.5, 250 mM Imidazole, 5% glycerol, 250 mM NaCl, 0.5 mM DTT), at the rate of 2 mL/min followed by isocratic wash with 100% buffer B. 5 mL fractions were collected over the course of gradient.

.

4.2.2.4 Tev-cleavable fusion tag removal

Fractions containing the nsP2SINV were pooled (based on the absorbance peak and SDS-PAGE analysis) and dialysed along with 1 mg/mL TEV protease against the dialysis buffer I (50 mM Tris pH 7.5, 5% glycerol, 0.5 mM DTT). Sample from fractions were run on 12% SDS-PAGE gel so as to understand the protein purity. The dialysed protein was then reloaded on the equilibrated His Trap Ni-NTA resin and the 6xhistidine tag cleaved protein is eluted in the flow through. The column is then washed with increasing the gradient with buffer B from 0% to

100%. Then cleaved protein and other fractions which are collected in the gradient are then loaded on the 12% SDS-PAGE so as to observe the cleavage.

4.2.2.5 Size exclusion chromatography

The cleaved protein is then concentrated using Amicon (10 kDa cut-off) till 1 mL volume and loaded on pre-equilibrated superdex 75 16/60 (50 mM Tris pH 7.5, 5% Glycerol, 0.5 mM DTT) column. Protein eluted at the desired column volume was then concentrated using Amicon [10 kDa cut-off], only after confirmation by loading on the SDS-PAGE. Protein was concentrated to around 10 mg/mL. It was critically important to keep fresh reducing agent (5 mM DTT or preferably 1 mM TCEP) present in the final size exclusion buffer.

4.2.2.6 Crystallization of nsP2SINV

The crystals of SINnsP2 were obtained by vapor diffusion method in 96-well sitting-drop plates (Hampton Research, USA) at 293 K. For crystallization experiment purified preparation of nsP2SINV was used at a concentration of 10 mg/mL in 50 mM Tris pH 7.0 and 100 mM NaCl. For the initial crystallization screening, small drops were prepared by mixing 1 µL of protein solution with the same volume of well solution and were equilibrated against 50 µL of well solution. Initial crystallization conditions were screened by using all the available screening kits from Hampton Research, USA.

4.3. Results and discussion

4.3.1. nsP2 sequence analysis and construct selection

Alphavirus nsP2 cDNA sequence was available from SINV, AURA, CHIKVRRV, SFV, and VEEV (see 4.2.1). A sequence alignment of full-length alphavirus nsP2 revealed good sequence identity throughout the polypeptide (Figure 4.2). The N-terminal ~460 amino acids of alphavirus nsP2 possessed significantly greater sequence identity (65-81% pair-wise sequence identity) than the C-terminal ~350 amino acids (Figure 4.2). The N-terminal ~460 amino acids, or the "helicase domain", contain remarkably few insertions among the 5 alphavirus sequences; only a single three amino acid insertion is found within SINV and AURA at position 299-301 (Figure 4.2). The C-terminal ~350 amino acids, or the "protease domain", contains a single six amino acid insertion within SINV and AURA at position 549-555 and a number of 1 to 2 amino acid insertions in the other viral sequences.

Disorder prediction is a powerful predictive tool for identifying potential sub domains by protein sequence analysis in the absence of other biological and biochemical information. An averaged FoldIndex disorder prediction based on a multiple sequence alignment correlated well with reports that nsP2 is a multi-domain protein (Figure 4.3). With proteins dissected from polyproteins that I or others within the laboratory have worked on, the final optimized construct correlated extremely well with the FoldIndex algorithm disorder prediction (Aura capsid protease [31], alphavirus nsP4 catalytic domain [32]).

Keeping this in mind, three different constructs of nsP2SINV were finalized; one complete C terminal nsP2 region comprising of protease and methyltransferase domain 472-801, two constructs comprising only the C terminal protease domain 472-603, 472-609. Oligonucleotides for these regions were designed on the basis of the nsP2SINV sequence available in the NCBI database (Table 4.1). It was specifically kept in the mind to avoid the hydrophobic residues at the N and C terminal ends of the construct so as to obtain a soluble heterologously expressed nsP2SINV protein.

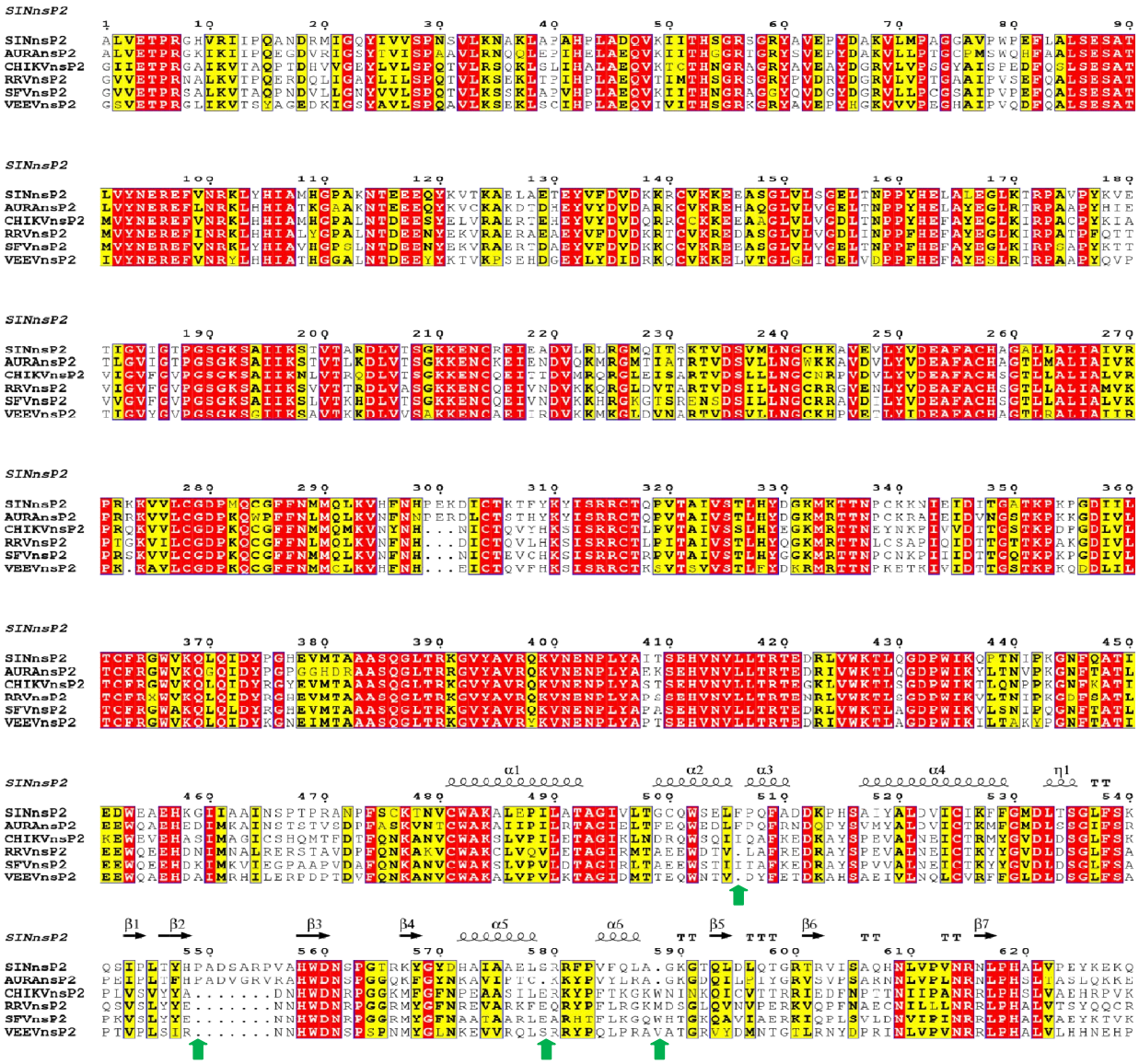


Figure 4.2 Multiple sequence alignment of nsP2 from SINV, AURA, CHIKV, SFV, RRV, and VEEV. Alignment was generated by ClustalW and reformatted by ESPript. Different amino acid insertions in the viruses in the protease domain are marked by the green arrow.

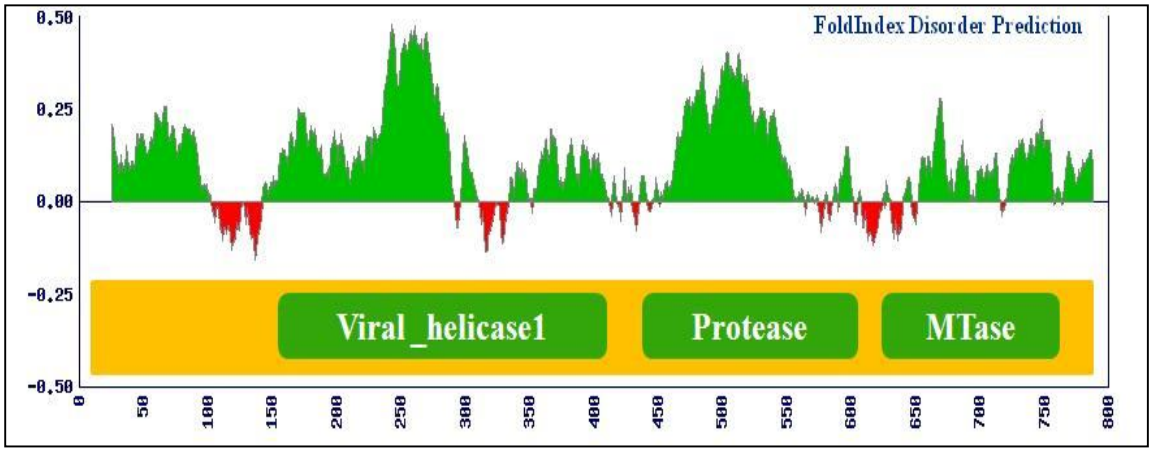


Figure 4.3 Alphavirus nsP2 domain organization. A multiple-sequence-alignment averaged FoldIndex disorder prediction as produced from Imatrix is shown (green is predicted as "ordered" and red is predicted as "disordered").

4.3.2. Cloning and Expression of different constructs of nsP2SINV

To get the recombinant soluble nsP2SINV protein, the gene encoding nsP2 C terminal protease was cloned into pET-28c-TEV, an *E.coli* based expression vector with an engineered N-terminal 6xhis tag and TEV protease cleavage site. The genomic cDNA of SINV in pToto64 was used for the PCR amplification of nsP2 protease gene by gene specific primers (Table 4.1). Analysis of PCR product on 1% agarose gel electrophoresis confirmed the presence of a single amplification band of desired (approx. 400bp, 450bp and 990bp) size, which almost matched to the gene size (Figure 4.4 A, B). The forward and reverse primers with restriction endonuclease sites of *NdeI* and *XhoI* were used for the PCR amplification. The vector DNA and PCR product were double digested with *NdeI* and *XhoI* restriction enzymes and T4 DNA ligase was used to ligate digested PCR product into digested pET-28c-TEV vector. The ligated product was successfully transformed in to *E. coli* DH5α cells. The plasmid DNA isolated from transformed cells was subjected to double digestion with *NdeI* and *XhoI* resulted yield the fragment of desired nsP2 protease gene length (approx. 400bp, 450bp and 990bp) (Figure 4.4 C, D). The automated DNA sequencing further confirmed the specificity of cloned DNA fragment. In order to optimize the

expression of these constructs, the recombinant vector was transformed into *E. coli Rossetta* (DE3) cells. The small scale expression of nsP2SINV constructs was analyzed at different I.P.T.G concentrations (0.1 mM to 1 mM) and different temperatures (4- 37°C). But no ideal condition was observed which gave us soluble protein [Figure 4.5 A-D]. None of the constructs were soluble at the above given conditions. So, we tried the addition of other components known to be contributing in the increase in solubility of heterologously expressed proteins. But addition of glycerol, phosphate salts, amino acids like glutamic acid and even sugar did not improve the solubility of the proteins. Finally we switched to different enriched media, and in one such media i.e auto-induction media we found solace. Although only one construct i.e full length C terminal nsP2 protease (472-801) was optimized to be soluble. The optimum expression and solubility was observed with 18 °C temperature with expression in auto-induction media.(Figure 4.5E).

Table 4.1 Oligonucleotides for the cloning of nsP2SINV constructs

Gene name	Insertion Vector	Primer direction	Primer sequence	Incorporated restriction enzyme site
nsP2SINV (472-801)	**pET28c**	**F**	**GATTCTCATATGAATCCGTTCAGCTGCAAGACC**	***NdeI***
		R	**GATTCTCTCGAGTTATGTACCCTCATACACGGACGA**	***XhoI***
nsP2SINV (472-615)	**pET28c**	**F**	**GATTCTCATATGAATCCGTTCAGCTGCAAGACCAACG**	***NdeI***
		R	**GATTCTCTCGAGTTAGCGGTTCACCGGGACCAGGTTAT**	***XhoI***
nsP2SINV (472-603)	**pET28c**	**F**	**GATTCTCATATGAATCCGTTCAGCTGCAAGACC**	***NdeI***
		R	**GATTCTCTCGAGTTAAACTCTGGTTCTCCCCGTCT**	***XhoI***

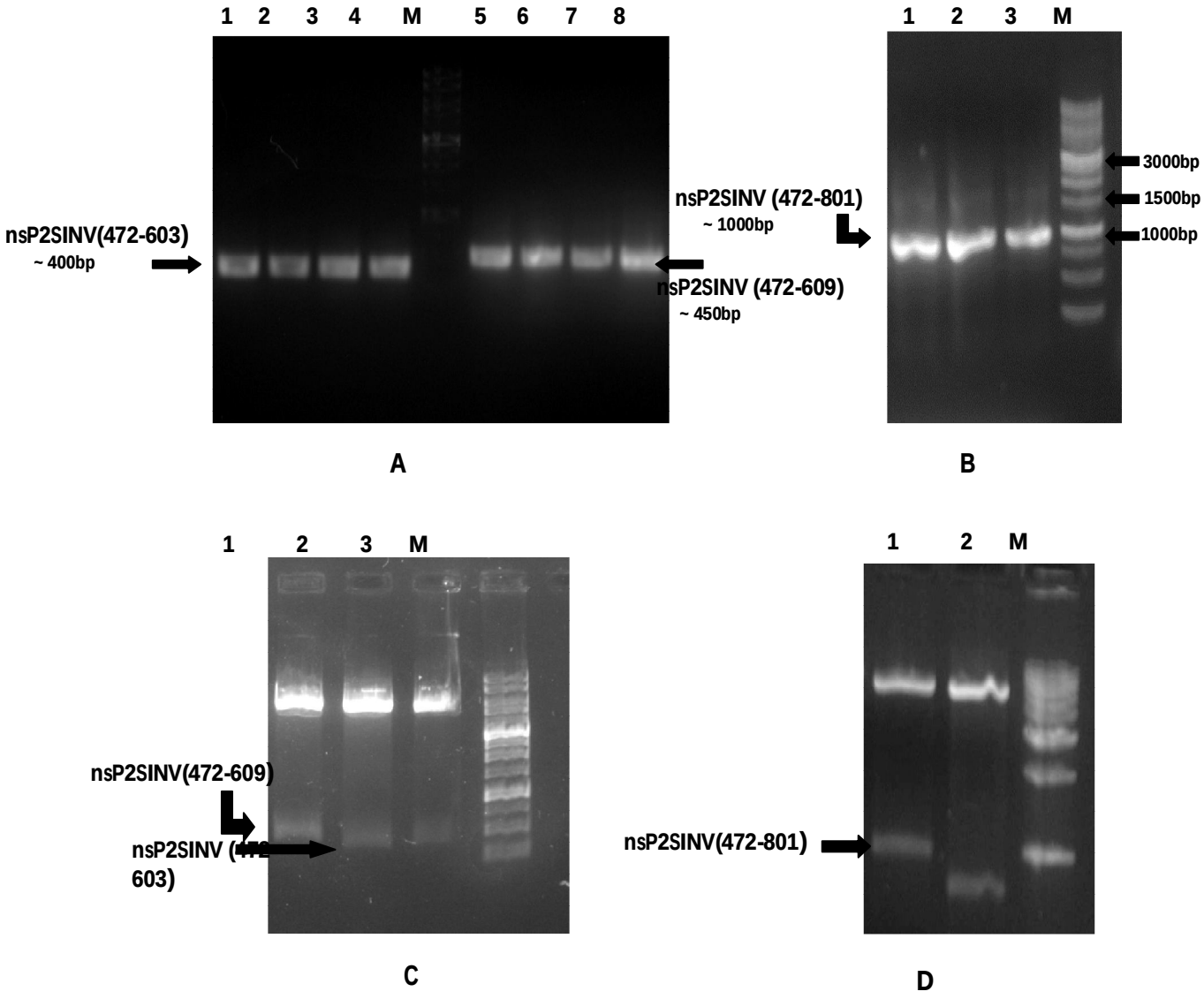


Figure 4.4 Amplification and cloning of nsP2SINV gene. **Agarose gel electrophoresis of (A)** PCR amplified products. lane M: 1 kb DNA marker; lane 1-4: PCR product (nsP2SINV (472-603)) amplified from cDNA with nsP2 specific primers. lane 5-8: PCR product (nsP2SINV 472-609)) amplified from cDNA with nsP2 specific primers **(B)** PCR amplified products. lane M: 1 kb DNA marker; lane 1-3: PCR product (nsP2SINV (472-801)). **(C)** Recombinant plasmids digested with *Nde*I and *Xho*I. lane M: 1 kb DNA marker; lane 1: pET-28c- nsP2SINV 472-609, lane 2, 3: pET-28c- nsP2SINV 472-603. **(D)** lane M: 1 kb DNA marker; lane 2: pET-28c-nsP2SINV (472-801).

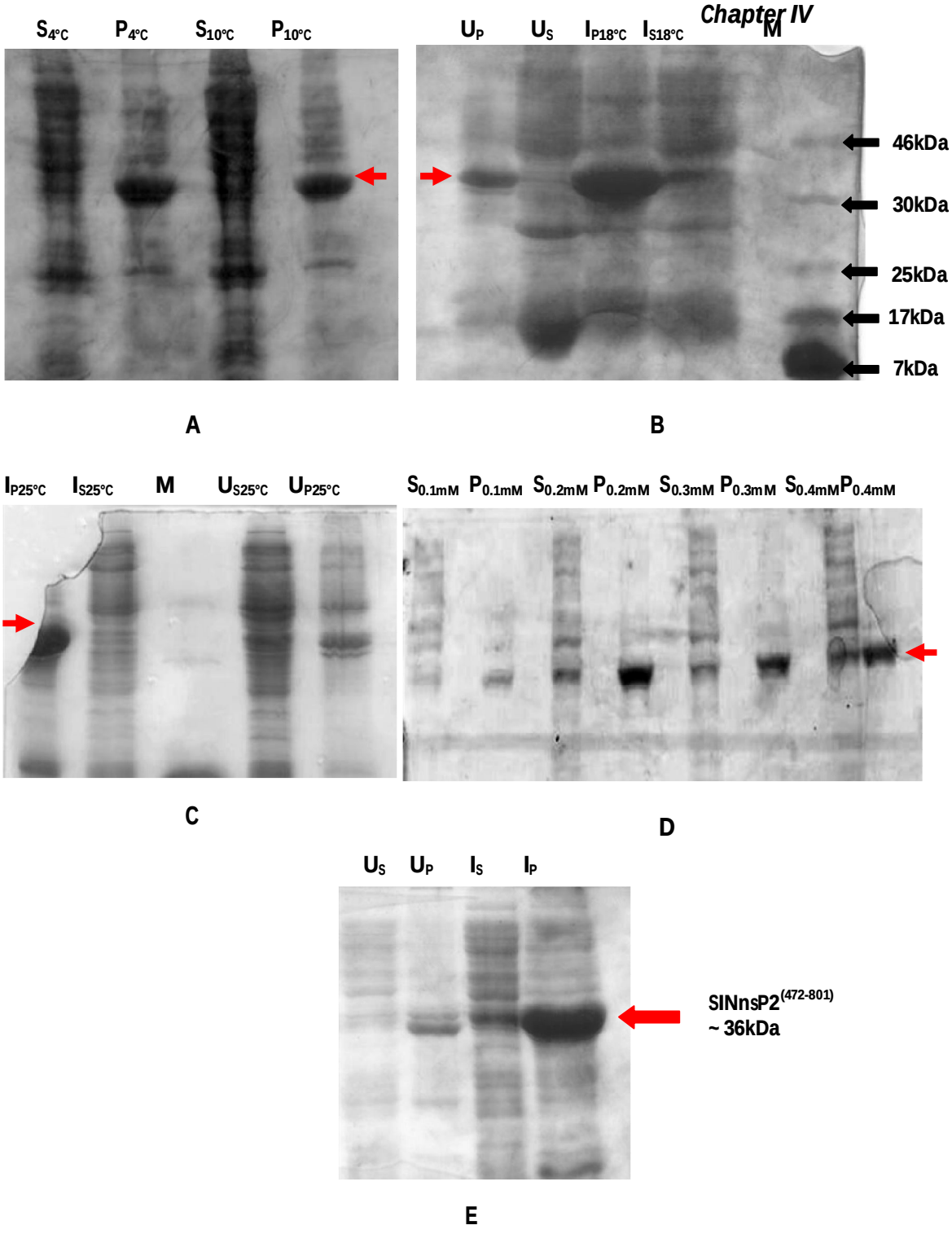
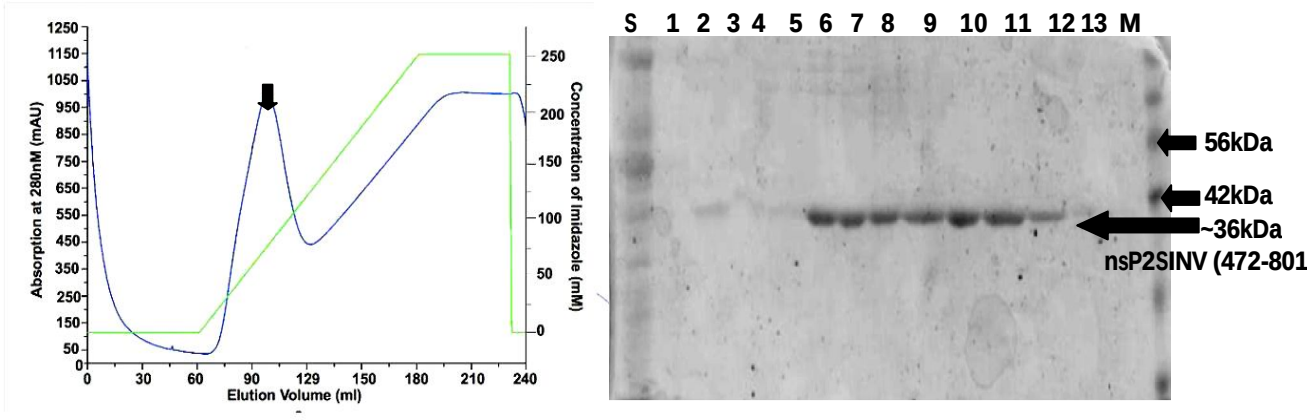


Figure 4.5 Optimized expression of nsP2SINV (472-801) construct (A-C). Expression at different temperature conditions. **S** is the supernatant; **P** is the pellet at different temperatures. M is NEB prestained broad range protein ladder (7-175 kDa). (D) Expression at different molar

concentrations of IPTG (E) Expression in auto-induction media. Red color arrow points towards the expressed ~36 kDa nsP2SINV (472-801) protease.

4.3.3 Purification of nsP2SINV (472-801)

pET-28cTEV-nsP2SINV (472-801) vector was transformed into *E. coli Rossetta* (DE3) cells using calcium chloride based heat shock method. Growth of cells in auto-induction media resulted into high level expression of nsP2SINV (472-801) protein (~25 mg/L) with N-terminal 6xhis tag. The protein was purified with the combination of two chromatography steps, immobilized metal affinity chromatography (IMAC) based affinity (Histrap Ni-NTA agarose) and gel filtration (Superdex-75).Pure protein was obtained by the affinity chromatography which was treated with purified TEV protease. Since there was a TEV protease cleavage site between the 6xhis tag and the protein, it resulted in the cleavage of the 6xhis tag. The treated protein was reloaded on the Histrap column, and most of the cleaved protein was obtained in the flow through. The cleaved protein was loaded on the gel filtration superdex 75 16/60 column for further purification. nsP2SINV (472-801) was purified up to more than 95% purity and homogeneity (Figure 4.6). Single preparation from 1 liter of induced culture typically yields 20-25 mg of nsP2SINV (472-801) protein with near 95% homogeneity. On superdex-75 16/60 prep grade gel filtration column, nsP2SINV (472-801) eluted as single sharp peak at volume which corresponds with its molecular mass of around 36 kDa (Figure 4.5 C). The gel filtration analysis suggests that nsP2SINV (472-801) remains in a monomeric state in aqueous solution. Since the protein was quite unstable and would aggregate very fast, thus DTT was added in the elution buffer and other purification buffers.



A B

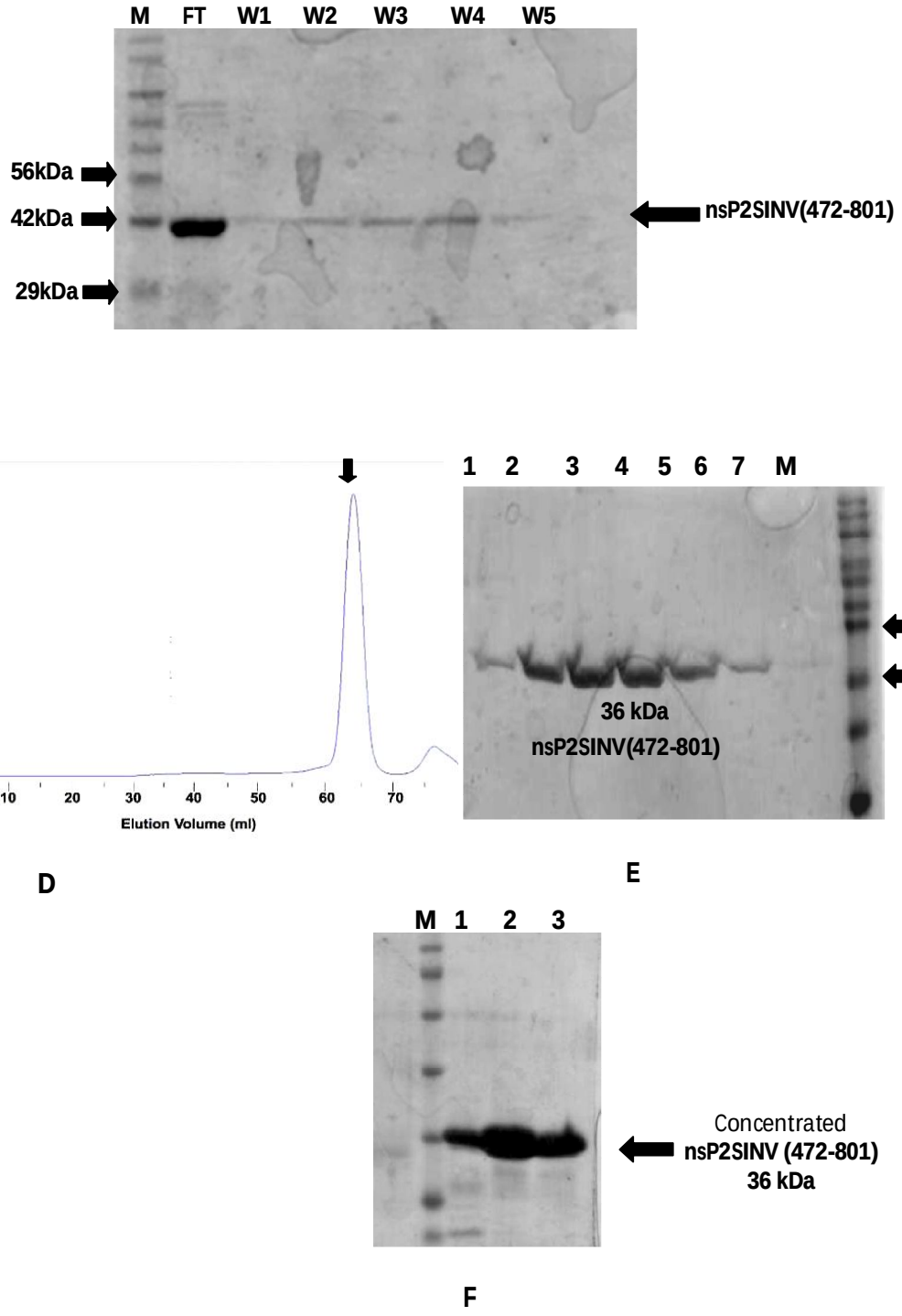


Figure 4.6 Optimized purification of recombinant nsP2SINV(472-801) (**A)** Chromatogram obtained from an extract of culture of E. coli Rossetta (DE3) cells containing the pET28c-TEV-nsP2SINV (472-801) plasmid applied to a Ni-NTA affinity chromatography column with peak corresponding to fractions containing nsP2SINV (472-801) indicated by an arrow. (**B)** Analysis of the affinity purification by SDSPAGE (12%) stained with Coomassie Blue. Soluble extract of

E. coli culture (S) and eluted fractions (lanes 5-11). **(C)** Reverse NTA of the TEV protease cleaved nsP2SINV (472-801) with the cleaved protein majorly coming in Flow though (FT) followed by wash fractions (W1-W5). **(D)** Chromatogram obtained after the TEV cleaved protein was concentrated and applied to a superdex-75 16/60 gel filtration column, respectively **(E)** Protein fractions from gel filtration chromatography. M is the protein ladder flow. **(F)** Concentrated nsP2SINV (472-801) protein used for crystallography.

4.3.4. Crystallization of nsP2SINV (472-801)

Commercial sparse matrix screens from Hampton research, USA were utilized for the initial crystal screening. Each screen had 96 different solutions varying in precipitant, salt and additives combination. Different buffers were used to screen a wide pH range (4.0-10.0). The concentration of nsP2SINV (472-801) was kept at 8 mg/mL for the initial screening. In the case of crystal screen conditions SINnsP2(472-801) crystallized initially as small needle like crystals with were observed in conditions consisting 0.2M Lithium sulfate monohydrate, 0.1M TRIS hydrochloride (pH 8.5), 30%w/v Polyethylene glycol 4,000 and 0.2 M Ammonium sulfate, 0.1M Sodium acetate trihydrate (pH 6.5), 30% Polyethylene glycol monomethyl ether 2,000.

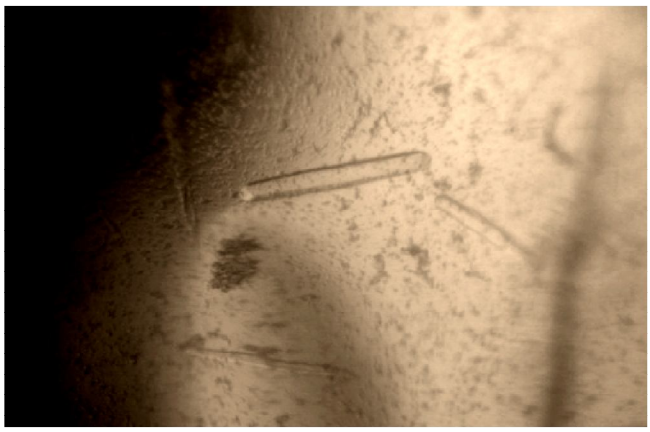


Figure 4.7 Crystals of nsP2SINV (472-801) obtained by the hanging-drop vapor-diffusion method using 0.2 M Ammonium sulfate, 0.1M Sodium acetate trihydrate (pH 6.5), 30% Polyethylene glycol monomethyl ether 2,000 as precipitants at 4°C.

The crystals were tried to improve manually playing around the conditions in which we got initial hits. The size of crystals however improved but they were not stable and would dissolve after some days of appearance.

4.4 Conclusion

Overall, the strategy of domain dissection as a tool for structure-function studies of alphavirus nsP2 proved to be a success, particularly when the immediate goal is to obtain soluble material. The results presented here demonstrated that construct selection is critical for the production of soluble protein from a difficult to study multi-domain, multifunction protein. Disorder prediction proved to be a powerful tool for the selection of constructs that would provide soluble protein expression in *E. coli* expression systems. In regards to nsP2 protease domain, it was clearly demonstrated that the most soluble protein produced is firstly were difficult to obtain in some cases and secondly in case of SINV it is a highly unstable one which needed constant presence of reducing agent like DTT for its stability.

However, we overcame all the difficulties and were successful in obtaining crystals of the nsP2SINV (472-801) protease domain. Although, there are miles to go to achieve which we have initially set as the target. The crystals need to be improved further and the analysis of the structural information which we would get might open new insights about the functioning of this domain in virus infectivity.

4.5 Bibliography

1. Kurkela, S., Manni, T., Vaheri, A., & Vapalahti, O. Causative agent of Pogosta disease isolated from blood and skin lesions." Emerging infectious diseases 10, no. 5 (2004): 889.
2. Hernandez, R., & Paredes, A. Sindbis virus as a model for studies of conformational changes in a metastable virus and the role of conformational changes in in vitro antibody neutralisation." Reviews in medical virology 19, no. 5 (2009): 257-272.
3. Gu, Y., Yang, Y., & Liu, Y. Imaging early steps of sindbis virus infection by total internal reflection fluorescence microscopy." Advances in virology 2011 (2011).
4. Calisher, C. H., Meurman, O., Brummer-Korvenkontio, M., Halonen, P. E., & Muth, D. J. Sensitive enzyme immunoassay for detecting immunoglobulin M antibodies to Sindbis virus and further evidence that Pogosta disease is caused by a western equine encephalitis complex virus." Journal of clinical microbiology 22, no. 4 (1985): 566-571.
5. Lundstrom, J. O., Vene, S., Saluzzo, J. F., & Niklasson, B. Antigenic comparison of Ockelbo virus isolates from Sweden and Russia with Sindbis virus isolates from Europe, Africa, and Australia: further evidence for variation among alphaviruses." The American journal of tropical medicine and hygiene 49, no. 5 (1993): 531-537.
6. Taylor RM, Hurlbut HS, Work TH, Kingston JR, Frothingham TE. Sindbis virus: a newly recognized arthropod-transmitted virus. Am J Trop Med Hyg 1955; 4: 844–62.
7. Woodall JP, Williams MC, Ellice JM. Sindbis infection in man. East Afr Virus Res Inst Rep 1962; 12: 17.
8. McIntosh BM, McGillivray GM, Dickinson DB, Malherbe H. Illness caused by Sindbis and West Nile viruses in South Africa. S Afr Med J 1964; 38: 291–4.
9. Tesh RB. Arthritides caused by mosquito-borne viruses. Ann Rev Med 1982; 33: 31–40.
10. Laine, M., R. Luukkainen, and A. Toivanen. "Sindbis viruses and other alphaviruses as cause of human arthritic disease." Journal of internal medicine 256, no. 6 (2004): 457-471.
11. Kurkela, Satu, Osmo Rätti, Eili Huhtamo, Nathalie Y. Uzcátegui, J. Pekka Nuorti, Juha Laakkonen, Tytti Manni, Pekka Helle, Antti Vaheri, and Olli Vapalahti. Sindbis virus infection in resident birds, migratory birds, and humans, Finland." Emerging infectious diseases 14, no. 1 (2008): 41.
12. Sane, Jussi, Sandra Guedes, Jukka Ollgren, Satu Kurkela, Peter Klemets, Olli Vapalahti, Eija Kela, Outi Lyytikäinen, and J. Pekka Nuorti. "Epidemic sindbis virus infection in Finland: a

population-based case-control study of risk factors." Journal of Infectious Diseases 204, no. 3 (2011): 459-466.

13. Brummer-Korvenkontio, Markus, Olli Vapalahti, P. Kuusisto, Pekka Saikku, Tytti Manni, Pentti Koskela, Tuire Nygren, Henrikki Brummer-Korvenkontio, and Antti Vaheri. "Epidemiology of Sindbis virus infections in Finland 1981–96: possible factors explaining a peculiar disease pattern." Epidemiology and infection 129, no. 02 (2002): 335-345.
14. Lundstro¨m JO, Lindstro¨m KM, Olsen B, Dufva R, Krakower DS. Prevalence of Sindbis virus neutralizing antibodies among Swedish passerines indicates that trushes are the main amplifying hosts. J Med Entomol 2001; 38: 289–97.
15. Norder H, Lundstro¨m JO., Kozuch O, Magnius LO. Genetic relatedness of Sindbis virus strains from Europe, Middle East and Africa. Virology 1996; 222: 440–5.
16. Francy DB, Jaenson TGT, Lundstro¨m JO et al. Ecologic studies of mosquitoes and birds as hosts of Ockelbo virus in Sweden and isolation of Inkoo and Batai viruses from mosquitoes. Am J Trop Med Hyg 1989; 41: 355–63.
17. Spradbrow PB. Arbovirus infections of domestic animals in Australia. Aust Vet J 1972; 48: 181–5.
18. Kozuch O, Labuda M, Nosek J. Isolation of Sindbis virus from the frog Rana ridibunda. Acta Virol 1978; 22: 78.
19. McIntosh BM, Jupp PG. Infections in sentinel pigeons by Sindbis and West Nile viruses in South Africa, with observations on Culex (Culex) univittatus (Diptera: Culicidae) attracted to these birds. J Med Entomol 1979; 16: 234–9.
20. Uggeldahl P-E. The view of a clinical dermatologist of the rash in Pogosta disease. Vopr Virusol 1985; 5: 636.
21. Skogh M, Espmark A° Ockelbo disease: epidemic arthritisexanthema syndrome in Sweden caused by Sindbis-virus like agent. Lancet 1982a; i: 795–6.
22. Brummer-Korvenkontio M, Kuusisto P. Onko Suomen la¨nsiosa sa¨a¨stynyt ‘Pogostalta’ (Has western Finland been spared the ‘Pogosta’?). Suom La¨a¨ka¨ril 1981; 32: 2606–7.
23. Lvov DK, Skvortsova TM, Kondrashina NG et al. Etiology of Karelian fever, a new arbovirus infection. Vopr Virusol 1982; 6: 690–2.
24. Niklasson B, Espmark A° , LeDuck JW et al. Association of a Sindbis-like virus with Ockelbo disease in Sweden. Am J Trop Med Hyg 1984; 33: 1212–17.

25. Laine M, Vainionpa¨a¨ R, Oksi J, Luukkainen R, Toivanen A. The prevalence of antibodies against Sindbis-related (Pogosta) virus in different parts of Finland. Rheumatology 2003; 42: 1–6.
26. Turunen M, Kuusisto P, Uggeldahl P-E, Toivanen A. Pogosta disease: clinical observations during an outbreak in the province of North Karelia, Finland. Br J Rheumatol 1998; 37: 1177–80.
27. Manni T, Kurkela S, Vaheri A, Vapalahti O. Diagnostics of Pogosta disease: antigenic properties and evaluation of Sindbis virus IgM and IgG enzyme immunoassays. Vector Borne Zoonotic Dis 2008;8:303-11.
28. Krauss, H., Schiefer, H. G., Weber, A., Slenczka, W., Appel, M., von Graevenitz, A., Enders, B., Zahner, H., & Isenberg, H. D. (2003). Viral Zoonoses. *Zoonoses: Infectious Diseases Transmissible from Animals to Humans* (3rd ed., pp. 15-16). Washington, D.C.: ASM Press.
29. McWilliam, Hamish, Weizhong Li, Mahmut Uludag, Silvano Squizzato, Young Mi Park, Nicola Buso, Andrew Peter Cowley, and Rodrigo Lopez. "Analysis tool web services from the EMBL-EBI." Nucleic acids research 41, no. W1 (2013): W597-W600.
30. Prilusky, Jaime, Clifford E. Felder, Tzviya Zeev-Ben-Mordehai, Edwin H. Rydberg, Orna Man, Jacques S. Beckmann, Israel Silman, and Joel L. Sussman. "FoldIndex©: a simple tool to predict whether a given protein sequence is intrinsically unfolded." Bioinformatics 21, no. 16 (2005): 3435-3438.
31. Aggarwal, Megha, Satya Tapas, Anjul Siwach, Pravindra Kumar, Richard J. Kuhn, and Shailly Tomar. "Crystal structure of aura virus capsid protease and its complex with dioxane: new insights into capsid-glycoprotein molecular contacts." (2012): e51288.
32. Tomar, Shailly, Richard W. Hardy, Janet L. Smith, and Richard J. Kuhn. "Catalytic core of alphavirus nonstructural protein nsP4 possesses terminal adenylyltransferase activity." Journal of virology 80, no. 20 (2006): 9962-9969.

Chapter V:

Structural analysis of nsP2 protease from Chikungunya and comparative analysis with other alphaviral non-structural proteases

Chapter V

5.1 Introduction

The nsP2 protease domain has multiple activities associated to it; regulating 26S sub-genomic RNA synthesis [1, 22], down regulation of the minus-strand RNA synthesis in the latter stages of the infection [23] and most importantly proteolytic processing of the non-structural polyprotein [24, 25]. The nsP2 proteases are cysteine proteases with a papain-like fold and belong to C9 family of clan CA which includes proteases having papain-like fold [5]. The members of this family have a characteristic catalytic core made up of one α-helix and three β-sheets [5, 6]. In contrast to other members of this family which have the catalytic cysteine at the α-helix and histidine at the β-loop of two separate domains and the active site quadrate at the intersection of them, in nsP2 proteases the active site is present in one single domain only [7]. The insights of the mechanism of the alphaviral nonstructural protease were proposed on the basis of the information available from nsP2VEEV protease (PDB: 2HWK) structure in conjunction with the reaction mechanism of a typical papain or papain-like proteases [8].

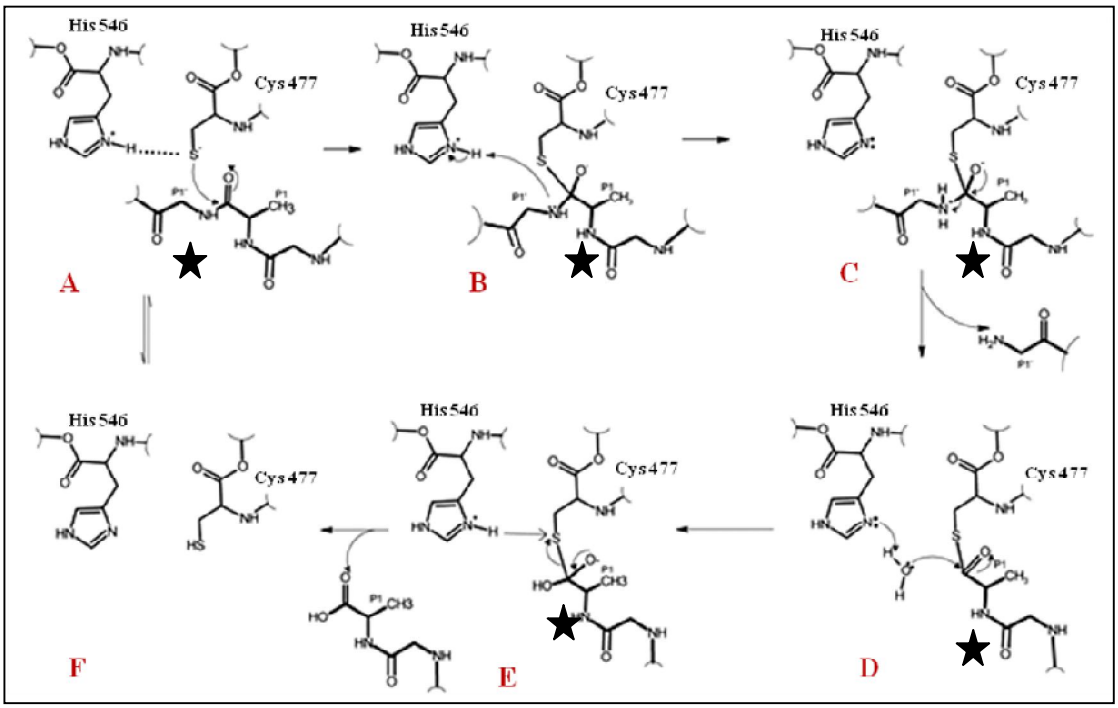


Figure 5.1 Proposed catalytic mechanism of nsP2VEEV protease. (A) The active site Cys477-His546 are in the ionized state represented by a thiolate/imidazolium ion pair. When substrate (indicated by ★) attaches at the active site, its P1 carbonyl carbon is attacked by the deprotonated thiol of cysteine. **(B)** Leading to the formation of negatively charged tetrahedral

intermediate with the P1' amine protonated by the imidazolium ion. **(C)** Which breaks down, releasing the C-terminal product and an acylated thiol. **(D)** And **(E)** which further undergoes hydrolysis releasing the N-terminal product and also free enzyme with active ionized catalytic dyad or the inactive uncharged dyad **(F)** Concept derived from ref**. [8, 9].**

In the absence of substrate, the histidine of the active site is not positioned in a way which is necessary for catalysis. A change in this rotameric state of histidine makes it catalytically active to form thiolate-imidazolium ion pair with active site cysteine. This conformation of the histidine is energetically favorable without even affecting the backbone conformation. In the presence of substrate this could be the rotameric form of this histidine which might assist the nucleophilic cysteine to attack on the scissile carbonyl carbon. pKa values predicted for the catalytic dyad residues (Cys 477= 5.1, His546=8.0 using PROPKA) on the basis of structural information have also supported this catalysis mechanism for the alphavirus non-structural protease [8]. Alphavirus non-structural proteases as discussed in chapter 1 are part of nsP2 C-terminal domain, playing pivotal role in the virus replication and infectivity in the host cell. When the virus infects the host cell, the genetic material is released in the host cell which is translated into either of the two types of non-structural polyproteins depending upon the alphavirus strain. This polyprotein is cleaved subsequently into proteins by the non-structural protease domain of the virus. Studies regarding the efficiency of these cleavages have demonstrated considerable differences at different sites; with preference order being nsp34 > nsp12 > nsp23 for both full length as well as isolated protease domain [10, 11, 12]. The nsP2 protease structure obtained from VEEV was able to impart some idea about the possible interactions between the protease domain and the substrate molecules but was not able to explain much about the basis for this preference order of the cleavage [7]. Some light has been shed about this preference order by the recently determined $nsp23^{pro\text{-}zbd}$ SINV crystal structure (PDB: 4GUA) [13] in which the arrangement of the four domains i.e protease, methyltransferase, macro, and zinc binding depicts the pre-cleavage form [13] (Figure 5.2). The carboxy terminus of this structure consists of the active site of the protease domain, and the positioning of the C-terminus of $nsP3^{zbd}$ points towards the possibility that the carboxyl region of the nsP3 might be able to cover this distance so to available for the *cis* cleavage by the protease. This large distance between the active site of the protease and the P2/3 site could be the one of the reason for the *trans* cleavage at the P2/3 site [13]. A comparative analysis of this structure

with the isolated structure of macro domains [14] and protease domains [7] of the other alphavirus members, reasons about the cleavage mechanism of the protease. It was observed that the protease and macro domain (nsP3 N- terminal region) have the most conserved structural features, whereas the MT-like domains showed some discrepancies [13]. The α-helix and β-sheet structure of the region superimpose well but the β-α loop regions are segregated by a distance of around 5Å, might be because of the interactions with the nsP3 linker region [13].

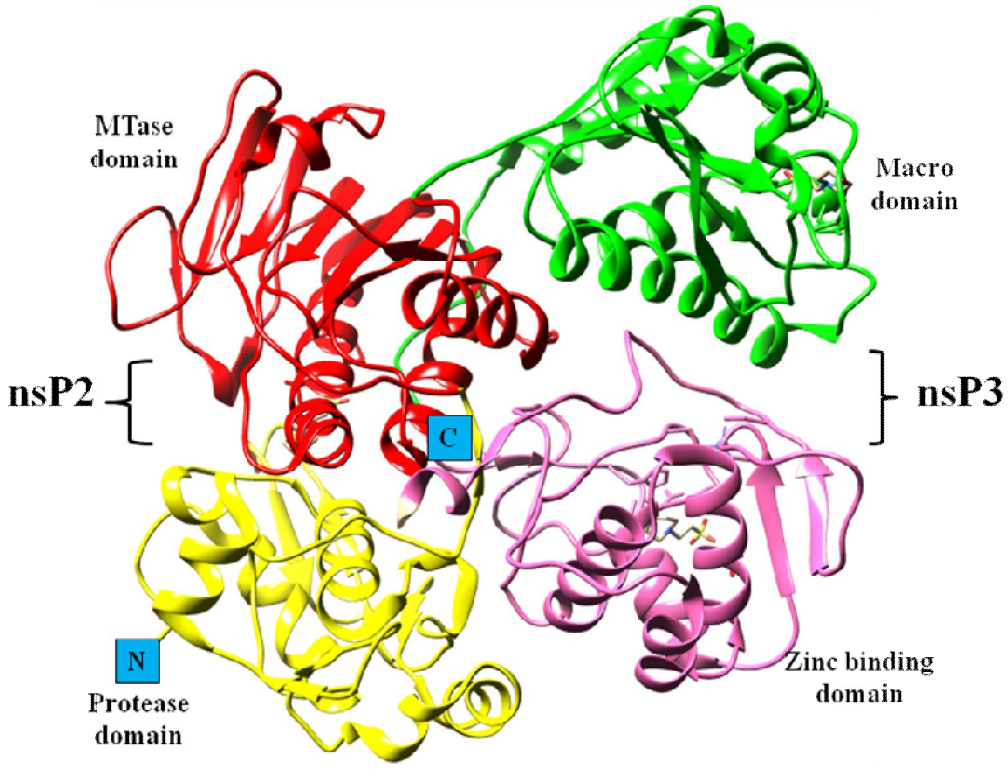


Figure 5.2 nsP23$^{pro\text{-}zbd}$SIN structure. Domains of nsP2 and nsP3: protease domain (yellow) and methyltransferase-like (red) are part of nsP2 C terminal region and the macro domain (green) and zinc-binding (pink) domain forms N terminal region of nsP3 **[13].** (PDB: 4GUA). Concept derived from Ref. **[13].**

Mutational studies in nsP23^{zbd}SINV of the interacting residues P726 (nsP2) and L165 (nsP3) exhibited reduced RNA infectivity and inefficient P2/3 cleavage [13]. This L165 is part of linker region which connects marcro domain with the zinc binding domain of nsP3 and this linker region further extends to the nsP2 methyltransferase region to interact with Pro726 (Figure 5.3 A) [13]. The surface around P726 is found to be basic and hydrophobic which might get

stabilized with the acidic nature of the L165 preventing any flexibility in the MT-like domain region, thus checking the P2/3 cleavage by the protease (Figure 5.3B) [13].

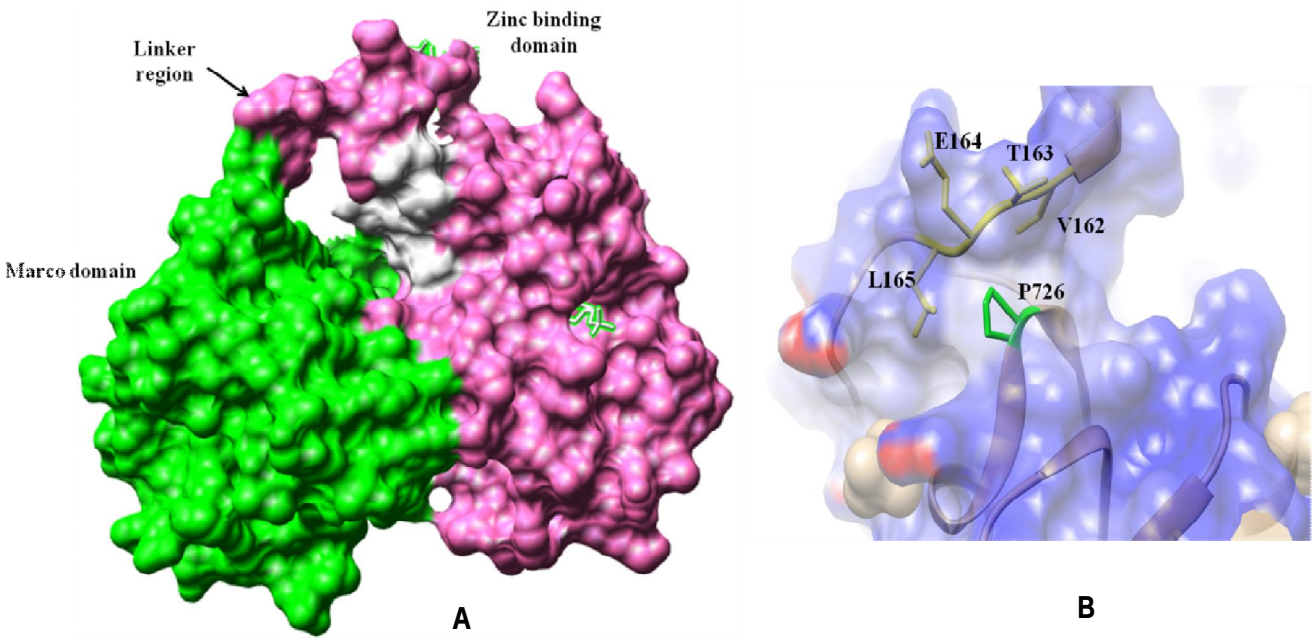


Figure 5.3 Interacting regions of nsP3 with nsP2 (A) Solvent-accessible region of nsP3. **(B)** Surface view of nsP2 P726 colored on the basis of electrostatic potential interacting with nsP3 linker region (visible in stick format). P726 is part of nsP2 whereas other labeled residues are from nsP3. nsP2/3 cleavage site is marked by an arrow. Concept derived from ref**.[13].**

Although the information available from these structures is not in symmetry and is vague in its own way but yet it does speaks volumes about the mechanism of the alphavirus non-structural protease, which is unique in its own way. We have tried to gather information for proteases in case of other members of this family, and were successful for one member. The nsP2 protease crystallized from CHIKV is quite similar yet have distinct structural features from the VEEV and SIN non structural protease structures. Here in this chapter we have described apo structure of nsP2 protease from CHIKV that can be utilized for structure-based drug discovery and novel drug development by means of co-crystallography and fragment based inhibitor screening. Also the structure would help in evaluating the significant features of the protein which helps in its distinguishing it from proteases available from other alphavirus members.

5.2 Materials and methods

5.2.1. Data collection

Diffraction size crystals were grown in 0.9 M sodium citrate tribasic dihydrate pH 6.0 at 4 °C with 1:1 protein: reservoir drop size using sitting drop method. Crystals were soaked for around 30 sec in the 1:4 70% glycerol: reservoir cryo-protectant solution before mounting. The diffraction data were collected at 100°K from a single large crystal using a a Bruker Microstar copper rotating anode X-ray generator (CuKα wavelength = 1.54 Å). The images were collected on MAR345dtb image plate detector. The crystal to detector distance was kept 200 mm and images were collected with exposure time of 10 min. and an oscillation width of 1^0 per image. The crystal diffracted maximum up to 2.60 Å resolution.

5.2.2. Structure refinement

For the structure solution, the reflections were indexed, integrated, and scaled using the HKL2000 program suite [27]. The nsP2CHIKV crystal belongs to space group *P212121*, the corresponding refinement statistics are shown in Table 2.3. The initial phases for nsP2CHIKV were obtained by molecular replacement with MOLREP [26] of the CCP4 suite. The protein coordinates of single subunit of nsP2CHIKV (PBD ID: 3TRK) that shares highest sequence and structure identity with our protein construct detected by the DALI server was used as suitable search template for molecular replacement. All ligands and waters coordinates were removed from the search template. This model provided sufficient phase estimates for subsequent model building and yielded a solution with one molecule per asymmetric unit. The reflections within the resolution range 74.59-2.59 Å were selected for refinement. The rigid body refinement was followed by iterative cycles of restrained atomic parameter refinement including TLS refinement with REFMAC5 [28] and PHENIX [29]. The repetitive cycles of model rebuilding based on σA-weighted 2Fo–Fc and Fo–Fc maps were performed by employing COOT [30]. The water molecules were added in the peaks contoured at 3σ in the Fo-Fc difference Fourier map which simultaneously satisfying density contoured at 1σ in the 2Fo-2Fc map.

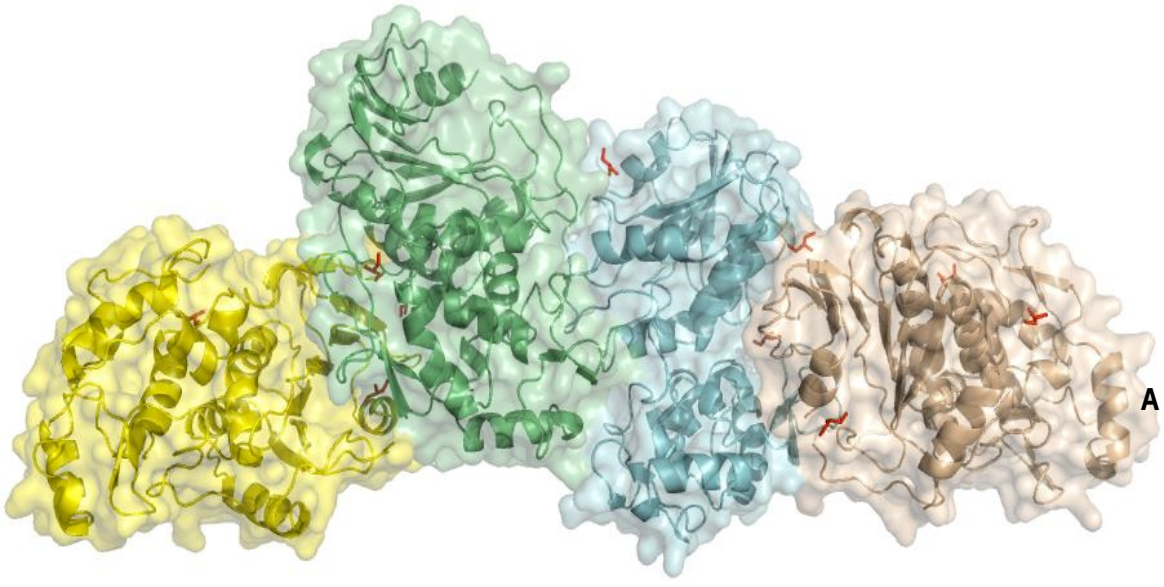
5.3. Result and discussion

5.3.1. Crystal structure of nsP2CHIKV protease

The structure of nsP2CHIKV protease was solved by X-ray crystallography using molecular replacement method [15], and was refined to R*free* of 25 %. The crystal belonged to P*212121* space group and diffracted up to 2.6 Å resolution (Table 2.3). The nsP2CHIKV protease was observed as a crystallographic homo-tetramer in this crystal form, having four monomers in the asymmetric unit as shown in (Figure 5.4 A).

The PDB was submitted for CHIKV nsP2 protease structure in the RCSB database with entry code **4ZTB**. This observation however is different with the results of size exclusion chromatography, native PAGE gel which showed nsP2CHIKV remains in monomeric state in aqueous solution (Figure 2.5, 2.6). The final model of nsP2CHIKV protease contained 1284 amino acid residues, 10 glycerol molecules and 262 water molecules.

The four chains interact with each other through different H bonds (shown as blue lines in Figure 5.4B) (Table 5.1) as well as some non-bonded contacts like salt bridges (range 2.9-3.9Å) (Figure 5.5) [16].



A

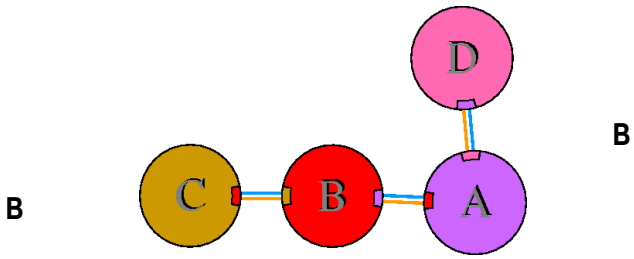


Figure 5.4 Schematic diagram of interactions between protein chains. **(A)** Surface view of all the chains of nsP2CHIKV interacting. Chain A (Teal), Chain B (Green), Chain C (Yellow) and Chain D (Brown). **(B)** Interacting chains are joined by colored lines, each representing a different type of interaction, as per the key above. The area of each circle is proportional to the surface area of the corresponding protein chain. (Figure made by PDBsum) **[17].**

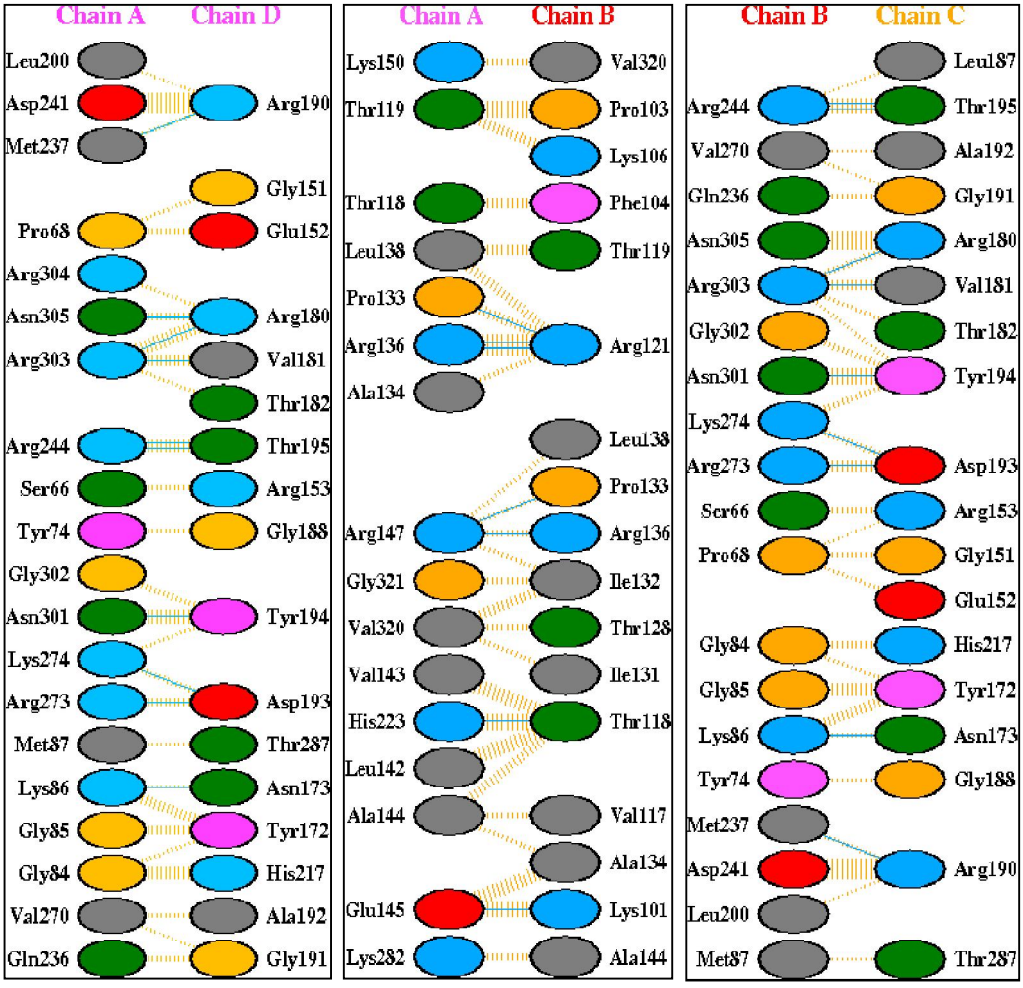


Figure 5.5 Chains of nsP2CHIKV interacting with each other. H bonds between the chains are shown as thin blue color lines. For non –bonded contacts, are shown as dashed lines and as width of the line increases with increase in the number of contacts. Residue colours: Positive

(Blue); negative (Red); Neutral (Green); Aliphatic (Grey); Aromatic (Purple, Pink); Pro&Gly (Brown); Cysteine (Yellow). (Using PDBsum server **[17]**.)

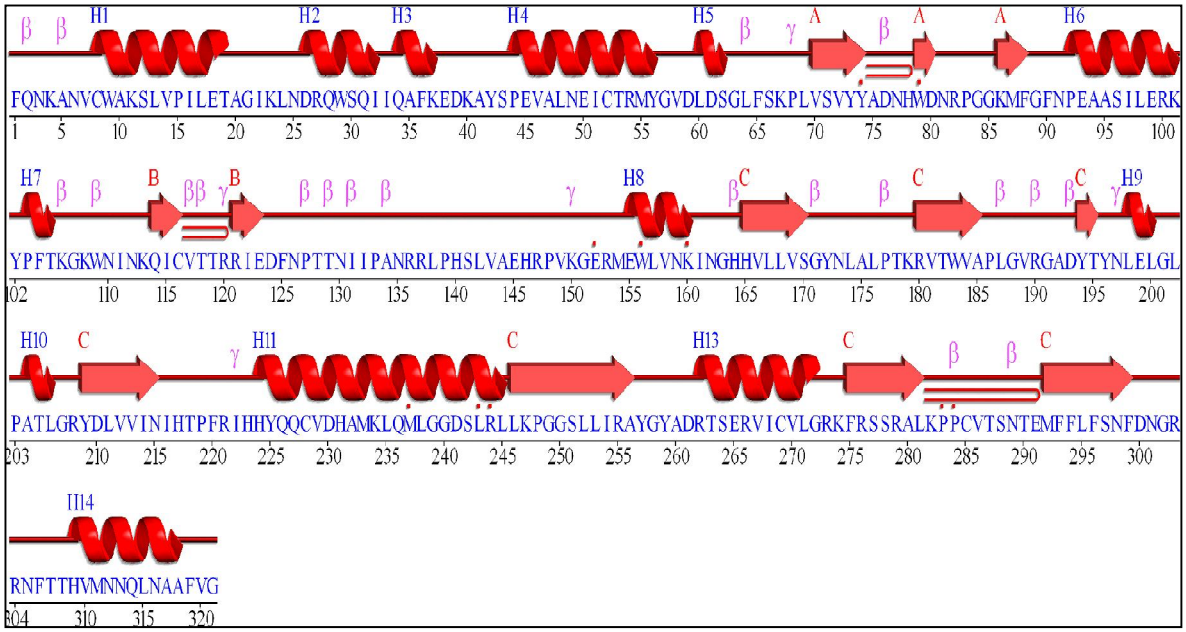


Figure 5.6 Overview of the secondary structure of nsP2CHIKV represented as wiring diagram prepared by PDBSum server [17]. α-helices labeled as H1 – H14 and β-strands by their sheets A, B and C. Residues involved in interaction with glycerol molecules are marked with red dots.

Table 5.1 Statistics for all the interfaces interactions in nsP2CHIKV protease

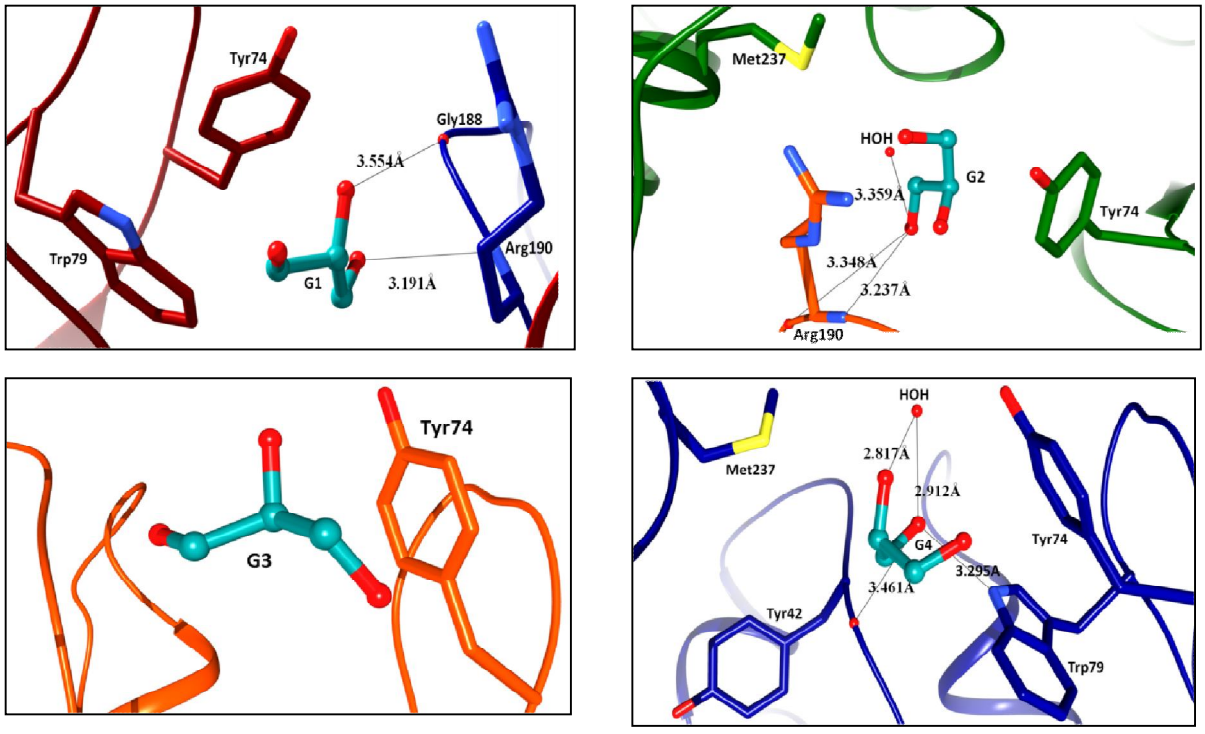
Chains	No. of interface Residues	Interface area (Å2)	No. of hydrogen Bonds	No. of non-bonded contacts
A }{ B	**16:17**	**902:915**	**7**	**48**
A }{ D	**20:17**	**894:916**	**10**	**85**
B }{ C	**19:18**	**879:904**	**9**	**89**

Table 5.2. Salt bridges in the nsP2CHIKV structure.

Amino Acid	Solvent Accessibility	Secondary structure	Salt bridge interactions (within chain)	Salt bridge interactions (inter-chain)
Lysine				
K40 A, B, C, D	Exposed	Loop	E38 - A, B, C, D	None
K113 A, B, C, D	Exposed	Loop	D124 - A, B, C, D	D124(C)- chain A None D124(A) - chain C None
K247 A, B, C, D	Exposed	β strand 9	D210 - A, B, C, D	None
K101 B	Exposed	Helix 6	E145	None
K106 C	Exposed	Loop	E99	None
Arginine				
R136 A, B, C, D	Buried	Loop	E265 (A, B, C, D)	None
R221 A, B, C, D	Exposed	Loop	D230(A, B, C, D)	None
R255 A, B, C, D	Buried	β strand 10	E291(A, B, C, D)	None
R266 A, B, D	Exposed	Helix 13	D61(A, B, D)	None
R273 A, B	Exposed	Loop	None	D193(D) - with chain A D193 (C) - with chain B
R304 A, B, C, D	Exposed	Loop	D300 (A, B, C, D)	None
R190 C, D	Exposed	Loop		D241(B) - C D241(A) - D
R247 C, D	Exposed	β strand 10	D210(C, D)	None
Histidine				
H165 A, B,C, D	Exposed	β strand 6	D210(A, B, C, D)	None

Majority of the Lysine and Arginine of all the chains are having different side chain conformations. This could be due to the fact that they are busy in stabilizing this tetrameric conformation of nsP2CHIKV via formation of the intra as well as inter-chain salt bridges with their negative charge counterparts. Different lysines, arginine forming salt bridges are described in the Table 5.2.

Unexpectedly the glycerol used in the purification and as a cryo-protectant was found to be bound at certain places in the final structure. Significance of this binding is not understood but it could be that this binding might be the implication of the flexibility of these regions which have allowed the access to the glycerol moiety. Since some of the glycerols have been observed to be bound near the active site regions like G1 (glycerol molecule 1), G3 (glycerol molecule 3) and G4 (glycerol molecule 4) (in Figure 5.7), they might be able to do so because of the flexibility of this region. But overall significance of this binding is still inconclusive.



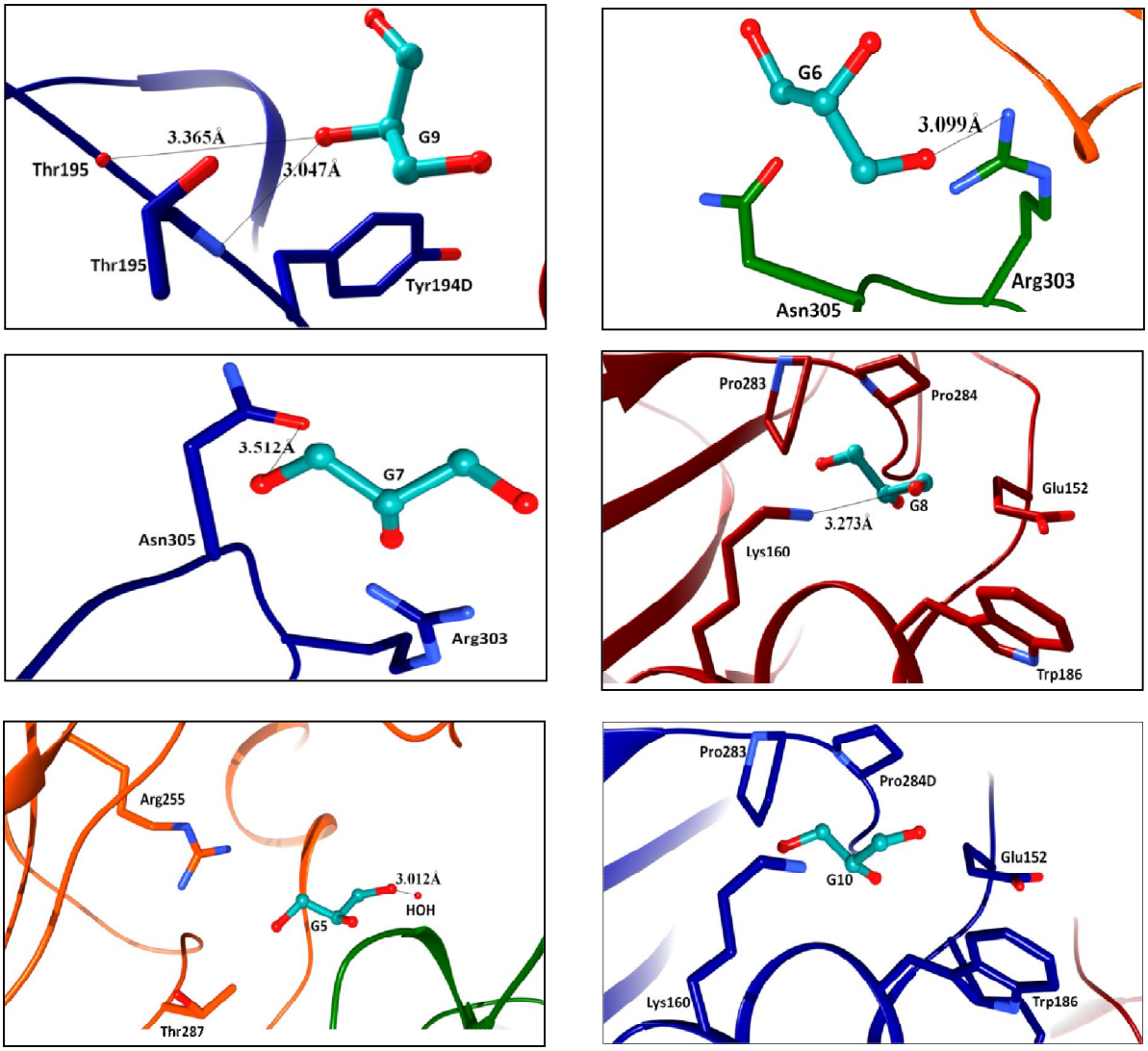


Figure 5.7. Glycerol molecules bound to the nsP2CHIKV chains. All the glycerols interacting with the residues are described here. Colour of the chains Chain A (Brick red), Chain B (Green), Chain C (Orange) and Chain D (Navy Blue).

The protease structure resolved at 2.56 Å consisting of residues from 1-321 (actual number in nsP2 sequence 471–791) is observed to adopt a compact bean-shaped fold (Figure 5.8). The secondary structure elements include 14 α- helices (Table 5.4) and three β-sheets (Table 5.3) (formed by 12 β-strands (Table 5.5)), topology arrangement is described in Figure 5.9. The protein structure clearly identifies two distinct domains. The N-terminal domain is moderately

small Phe1-Asn135 (Phe471-Asn605 in complete sequence of nsP2) from the C-terminal domain Arg136-Gly321 (Arg606-Gly791 in nsP2 complete sequence) and is seemingly helically dominant whereas the C-terminal domain consists of equal number of helices and β-sheets aligned horizontally to each other. The two domains are connected to each other via a random coil spanning from Asp124-Met154 (Asp594-Met624 in complete nsP2 sequence) and inter surface region is made up of helices and random coils. The N-terminal protease domain consists of seven helices surrounded by β hairpin on either side. This domain possesses both the residues of the catalytic dyad. The catalytic Cys8 (Cys478) is located at the N-terminus of helix α1 and the other catalytic residue His78 (His548) is located on the loop along with the third implicit active site residue Trp79 (Trp549) which resides on the β2 strand of sheet I. The substrate binding cleft is lined by the loop connecting β1-strand and β2 and the N-terminal region of helix α1 on one side and by β-strand 2 on the other side. The catalytic dyad and active site in case of nsP2CHIKV is similar to that of papain but the varaiation in tertiary structure is increases as we move away from active site. These variations in the features of nsP2 protease points towards the uniqueness, and hence novelty in the folding of the protease domain (Figure 5.10).

Table 5.3 The table gives the topology of the sheet using the nomenclature of Richardson (1981). Each connection between strands is defined by a number which represents the number of strands it crosses over in the sheet and in which direction, with an 'X' added for crossover connections.

Sheet	No. of strands	Type	Topology
A	**3**	**Antiparallel**	**1 -2X**
B	**2**	**Antiparallel**	**1**
C	**7**	**Mixed**	**1X 1X -3X -1X -2X 1**

Number of sheets in chain monomeric nsP2CHIKV protease: 3

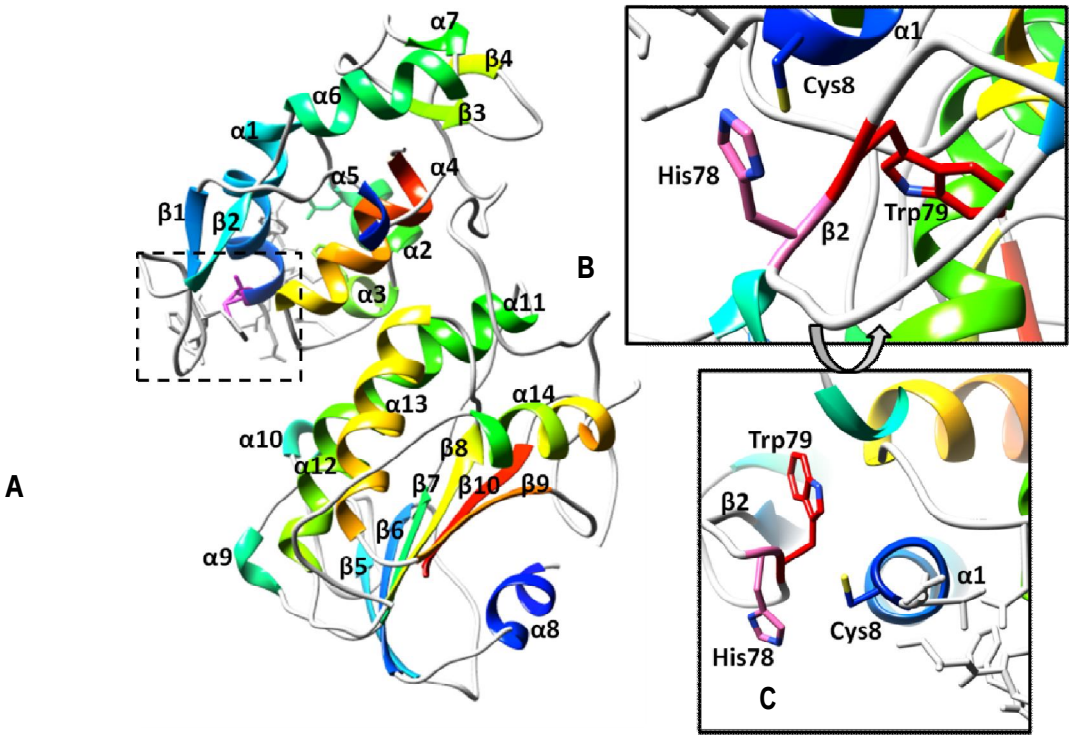
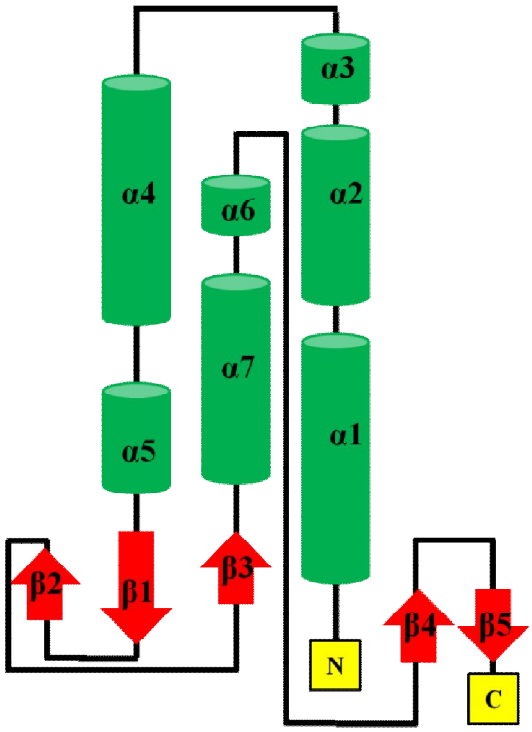


Figure 5.8 Cartoon representation of the model structure of the nsP2CHIKV (A) Secondary structure element labels are indicated and colored by secondary structure succession. The active site catalytic residues are shown as ball and sticks**. (B and C)** Cartoon representation of the first (helix α1) and the strand (β2) lids of CHIKVnsP2.



A

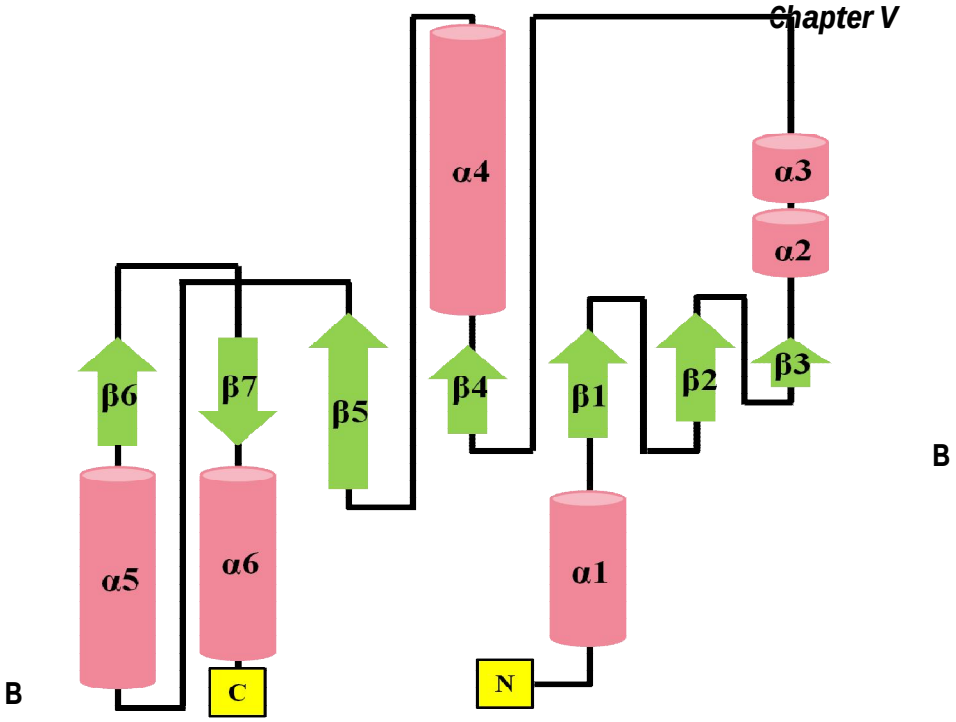
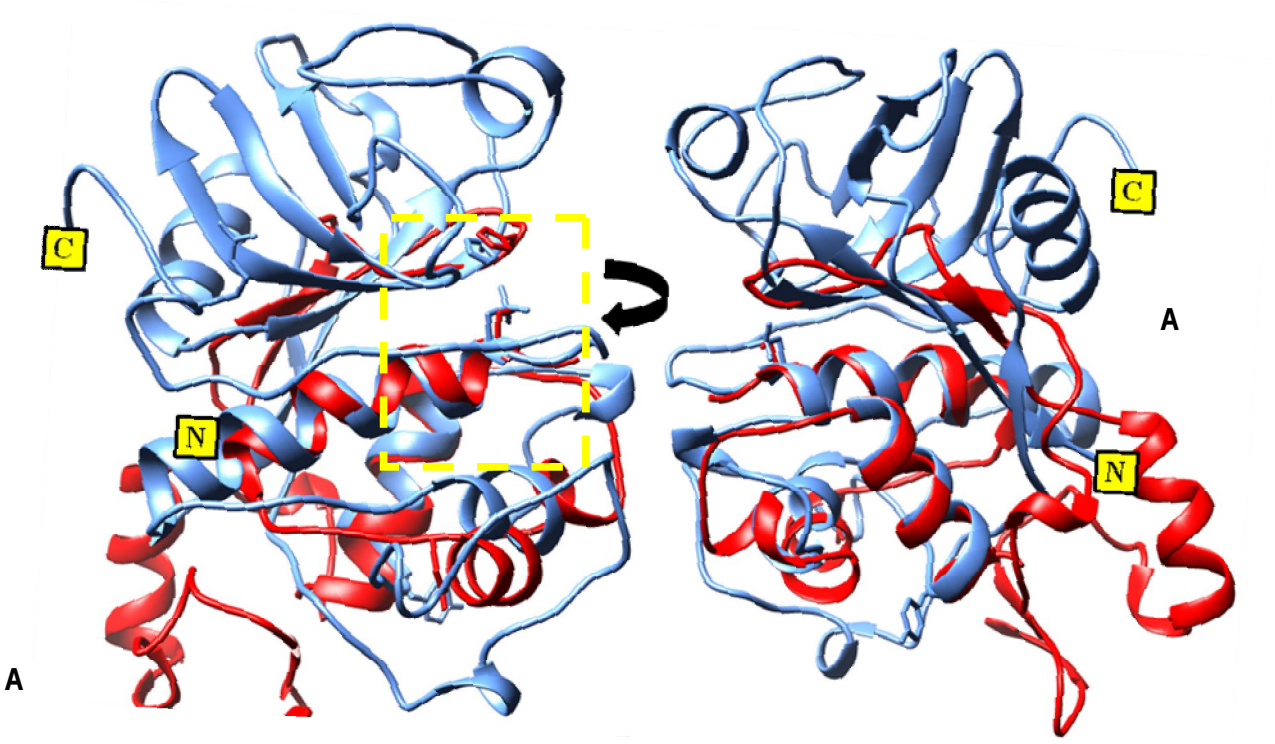


Figure 5.9 Topology representation of nsP2CHIKV protease monomer. The diagram illustrates how the twelve β-strands, represented by the block arrows, join up, side-by-side, to form the three β-sheets. The diagram also shows the relative locations of the fourteen α-helices, here represented by cylinders. (**A**) Describes the topology of N-terminal protease domain whereas topology of C-terminal methyltransferase domain is described in **(B).**



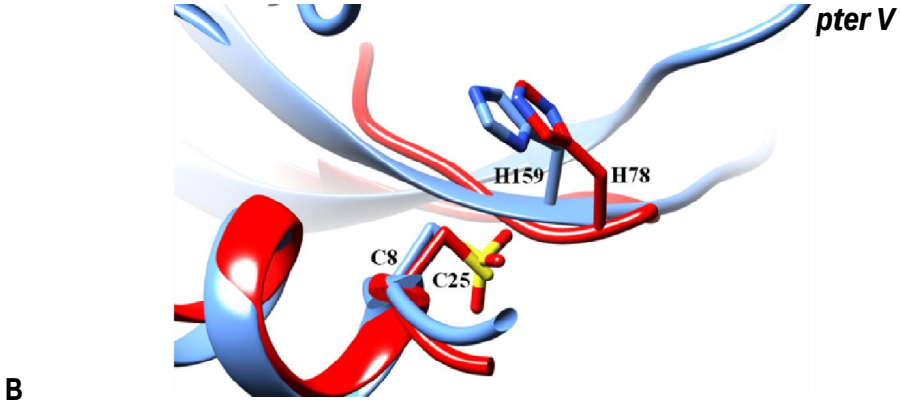


Figure 5.10 Comparison of nsP2 with Papain. (A) The N-terminal protease domain of nsP2 (color red) aligns well with papain (blue) near the active site region (marked by yellow lines). **(B)** And the active site residues C25, H158 (papain PDB: 9PAP) also are in symmetry with the active site residues of nsP2CHIKV (C8, H78).

Table 5.4 Description of helices in nsP2CHIKV structure

Number	Start	End	Number of residues	Sequence
1	Cys8	Thr19	12	CWAKSLVPILET
2	Asp26	Ile32	7	DRQWSQI
3	Gln34	Lys37	4	QAFK
4	Pro44	Tyr56	13	PEVALNEICTRMY
5	Leu60	Ser62	3	LDS
6	Pro92	Lys101	10	PEAASILERK
7	Pro103	Thr105	3	PFT
8	Glu155	Lys160	6	EWLVNK
9	Leu198	Leu200	3	LEL
10	Ala204	Leu206	3	ATL
11	His224	Asp241	18	HYQQCVDHAMKLQMLGGD
12	Ser242	Leu245	3	SLRL
13	Arg262	Gly272	11	RTSERVICVLG
14	His309	Ala318	10	HVMNNQLNAA

Table 5.5 Description of the β-strands for nsP2CHIKV structure.

Number	Start	End	Sheet	Number of residues	Sequence
1	Val70	Tyr74	A	5	VSVYY
2	Trp79	Asp80	A	2	WD
3	Lys86	Phe88	B	3	KMF
4	Gln114	Cys116	B	3	QIC
5	Arg121	Glu123	C	3	RIE
6	His165	Ser170	C	6	HVLLVS
7	Arg180	Ala185	C	6	RVTWVA
8	Tyr194	Thr195	C	2	YT
9	Tyr209	Asn215	C	7	YDLVVIN
10	Leu246	Ala256	C	11	LKPGGSLLIRA
11	Phe275	Leu281	C	7	FRSSRAL
12	Met292	Phe299	C	8	MFFLFSNF

Active site

The residues lining the active site of the nsP2 protease are not conserved expect few amino acids in the immediate vicinity of the catalytic residues. Classical cysteine protease derives either of the residues from two separate domains which come together at the interface of the domains to form the catalytic dyad. But in the nsP2 proteases these residues come from the same domain conferring that the nsP2 N-terminal protease domain is novel. Mutational studies along with the structural data confirm that Cys8 (Cys 478 in complete nsP2) and His78 (His 548 in complete nsP2 sequence) combined together forms the catalytic dyad of the nsP2 protease. And the third residue possibly playing role as the third element of the catalytic triad is assumed to be Trp79 (Trp 549 in complete nsP2 sequence) next to the catalytic histidine. This residue has been observed to be critical for binding with the substrate residues, thus playing role in the proteolytic activity of the protease (Figure 5.10) [18].

Substrate binding region

The substrate binding region could be described as a deep pronounced groove transversing the active site, contributing the residues towards binding of the peptide substrate in proximity to the catalytic dyad. The active site in CHIKV is observed as a closed tunnel in contrast to the nsP2VEEV (PDB ID: 2HWK) and nsP23^{prozbd}SIN (PDB ID:4GUA) alphavirus structures where the site is observed as an open channel. In the absence of any complex structure for the alphaviral non structural protease in-silico docking studies in case of nsP2VEEV protease provides insight about the active site region. The substrate binding residues Leu200 (Leu665), Met237 (Met707) and Asp241 (Asp711) of chain A and chain B were also observed to be interacting with Arg190 of chain C and chain D (Figure 5.6). Also this Leu200 is also observed to be interacting with N76 and contributing in the closed shape of active site in case of CHIKV nsP2 (Figure 5.12 C).

Docking of the Glu-Ala-Gly-Ala peptide substrate with the VEEV nsP2 described the locations for S1, S2, S3 and S4 sites on the protease. The binding site is a long deep grove dominated with basic amino acids formed by the interface residues of the nsP2protease N and C-terminal domains.

N-terminal domain predominantly occupies the substrate binding residues. Although C-terminal residues are not integral part of the catalytic machinery but they do provide pivotal residues for substrate binding. The entry into the active site seems to be a highly regulated process observed to be controlled by residues like Asp76, Asn77 (Asp546, Asn547 in complete nsP2 sequence) [8] residing towards the end of dynamic β-hairpin. These residues lining this region are observed to be having high temperature factors pointing towards the flexibility of this region [8].

In SINV antigenic members this region between residue Asn77 and His78 however has insertion of 7 extra amino acids (Figure 5.11). Substrate binding site is observed to accommodate upto 5 substrate residues in case of modeling studies with nsP2VEEV, and further added residue is observed to be positioned beyond this groove [8]. Different residues of nsP2CHIKV with which the P1, P2, P3, P4 and P1' residues of the substrate interact are described in Table 5.6. Most of the residues of these substrate binding sites are observed to be conserved in different alphavirus members.

Table 5.6 Substrate binding residues of nsP2CHIKV. Contacts of these residues with different cleavage sites are also provided.

Site nsP1/2	nsP2/3	nsP3/4	nsP2 residues (VEEV)	Contacts nsP1/2	nsP2/3	nsP3/4	nsP2 residues (CHIKV)
P1' Gly536	Ala1333	Tyr1863	Ala474		+	+	Ala475
			Asn475	+	+	+	Asn476
			Cys477*	+	+	+	Cys478*
			Lys480			+/-	Lys481
			His546*	+/-	+/-	+/-	His548*
P1 Ala535	Cys1332	Gly1862	Asn475	+	+	+	Asn476
			Cys477*	+	+/-	+	Cys478*
			Trp478	+/-			Trp479
			Ala509	+	+	+	Ala511
			Asn545	+/-		+/-	Asn547
			His546*	+	+	+	His548*
			Leu665	+	+	+	Leu668
P2 Gly534	Gly1331	Gly1861	Cys477*	+		+	Cys478*
			Trp478		+		Trp479
			Ala509	+	+	+	Ala511
			His510	+/-	+/-	+/-	His512
			Trp547*#	+	+	+	Trp549*#
P3 Ala533	Ala1330	Ala1860	His510	+	+	+	Trp512
			Ile542	+		+	Trp544
			Trp547	+	+	+	Trp549
			Ile698	+	+	+	Met703
			Met702	+	+	+	Met707
P4 Arg532	Arg1329	Arg1859	Ser511	+/-	+/-	+/-	Ser513
			Glu513	+/-			Glu515
			Trp547	+	+	+	Trp549
			Met702			+	Met707
			Lys706	+/-	+/-	+/-	Asp711

+ VDW contact, +/- Hydrogen bond, Residue in red are variable

Of all the residues involved in the binding of substrate peptides most significant contribution in defining the binding site is done by Trp79 (Trp549) which is conserved among all alphaviruses. In cysteine proteases which identify substrates with glycine motif at P2 position, they usually have bulky aromatic amino acids following the active site histidine [19, 20, 21]. These residues are the ones which generally define the S2 site of the protease. This describe they GSM strategy used by the alphavirus nsp2 protease for interaction between the active site tryptophan and P2 glycine of substrate [18].

C-terminal methyltransferase domain

The C-terminal domain extends from Arg136 (Arg606) to Gly321 (Gly791). The numbers of helices in this region are almost equal to that of number of β-strands in an arrangement in which the β-sheets are aligned parallel to helices. The arrangement of β-strands in nsP2CHIKV protease C-terminal region is similar to that in the methyltransferases. The topological arrangement in case of nsP2SINV and nsP2CHIKV is almost alike in this region of the protease. Although the C-terminal topological arrangement of nsP2CHIKV protease is similar to that of SINV but the active site region of the protein i.e N-terminal domain depicts differences in topology from both SINV and VEEV, which however share similarity in the sheet and helix arrangement with each other. Also the N-terminal region is extended by 4 residues in case of nsP2SINV protease but in case of CHIKV and VEEV nsP2 protease it concludes at the N135 (N605 in complete nsP2 sequence) and N1140.

A classical Rossman fold symmetry has an arrangement of β-strands arranged in the order 3-2-1-4-5-7-6 depicting a sheet sandwiched between alpha helices. This description is apt in case of the C-terminal region of nsP2 protease. This region is having significant similarity with the known methyltransferase structures. The strand region fits perfectly and aligns well with the tertiary structure of FtsJ methyltransferase fold [7]. But alignment is poor in the region involved in the interaction with SAM, the substrate for methyltransferases and critical residues involved for binding with the SAM are also absent in this domain. Sequence identity is also low for this region in alphaviruses. These all observations implicit this region to be a non-functional methyltransferase domain, which lacks the enzymatic methyltransferase activity in spite of its structural similarity with the known functional methyltransferases. However, studies of this region signify that this domain could be playing role as a scaffold for interactions with RNA [1, 2] or other viral non-structural proteins [1] and has importance in the regulation of the viral replication process [2].

Secondary Structure and Sequence Analysis

Sequence alignment of nsP2CHIKV protease with other alphavirus members shows limited conservation in the sequence. Only 18% of the residues of this region are conserved including the catalytic Cysteine and Histidine. And also the tryptophan which forms the part of the catalytic triad is also observed to be conserved in all the members (Figure 5.11). However despite of this low identity in the sequence the secondary structure arrangement and overall 3D configuration of these proteins is conserved in alphaviruses.

Structural insights into the catalytic mechanism

Comparison of the non structural proteases from different alphaviruses provides information about the probable structural changes which the protease might be undergoing during switching from one active state to the other. A closer examination of the structures between the identical regions by superposition investigates significant differences. The active site entry is regulated by an occluding loop like part of the active site which comprise of the active site histidine and tryptophan.

All the structure available for the non structural protease domain from the alphavirus members are in apo form [7, 13] and not a single structure is there where any protein has been observed to be interacting with its substrate peptide. In case of nsP2SINV structure which is available in its pro cleavage form it became clear that the 2/3 site is quite distant from the active site region. So in this case also we could not describe the exact mechanism of the proteolytic activity of this enzyme. Comparative analysis of structure nsP2VEEV (PDB -2HWK), nsP2SINV (PDB-4GUA) with this available nsP2CHIKV structure however generated information about the regions which are variable and thus it could be termed as the flexible regions in the nsP2 protease structures. Individually comparing the different regions showed changes which might have taken place during the course of time in the enzyme in the presence of different crystallizing conditions. In 2HWK three citrate molecules were found to be bound in the main protein chain out of which one citrate was bound near the active site. In case of nsP2SINV (PDB-4GUA), some sulphate ions were found to be near the active site region.

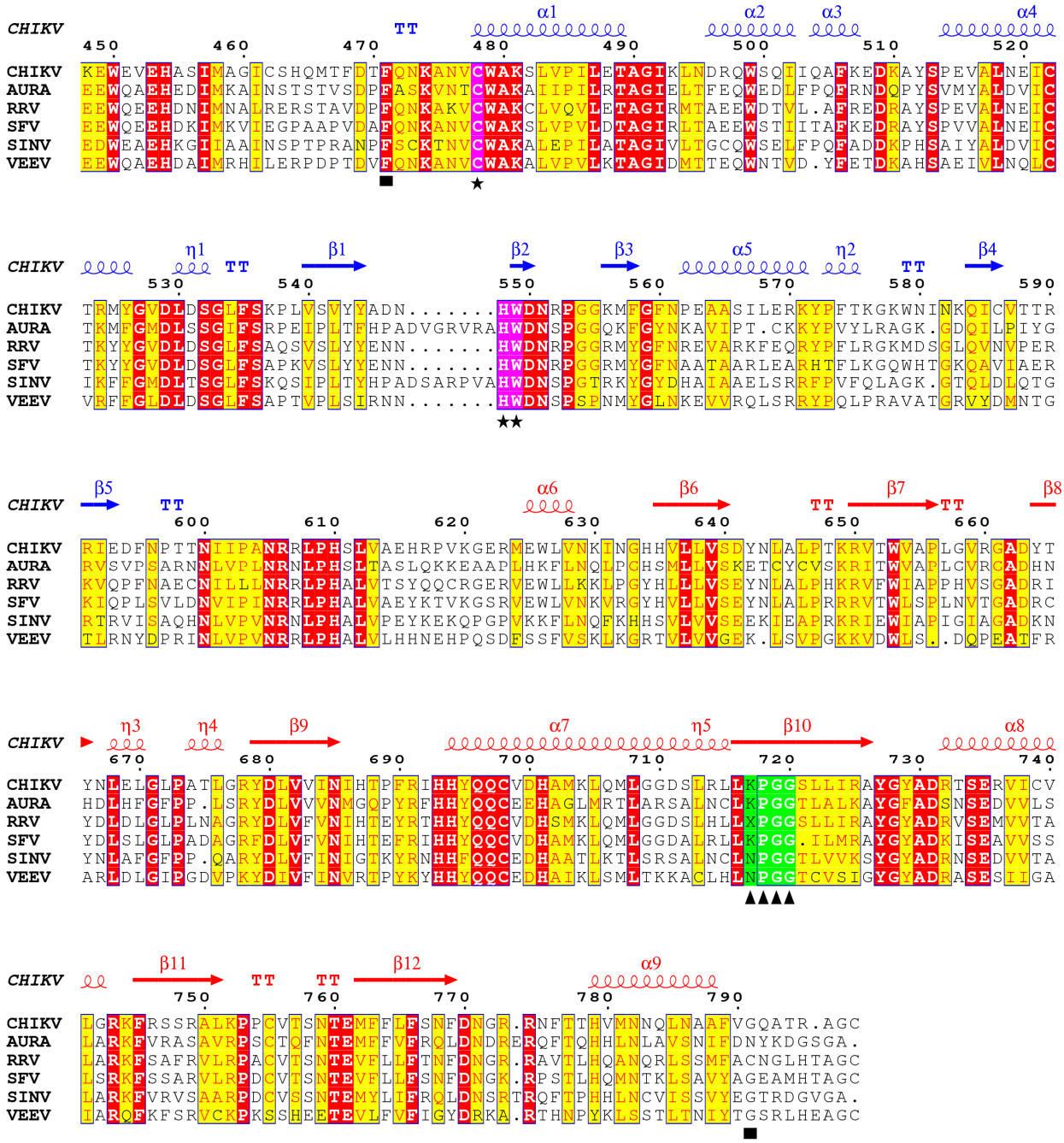
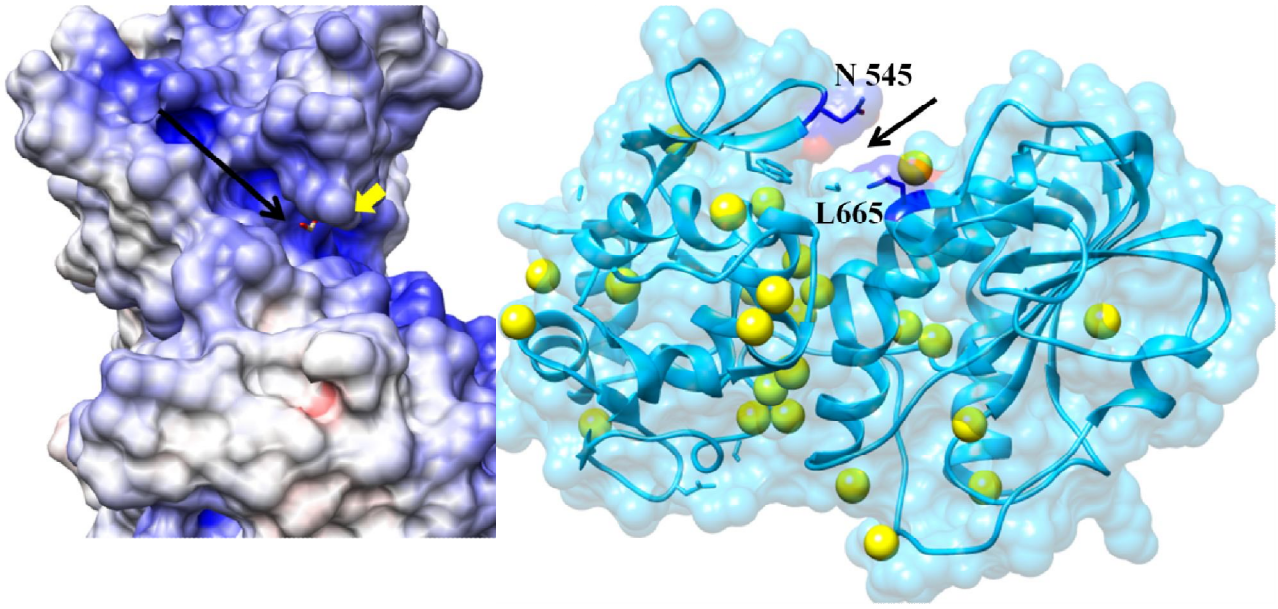


Figure 5.11 Sequence analysis of some important members of alphavirus family. Active site residues are colored Pink and marked with ★ , region cloned and crystallized for this study is defined with ■ symbols. A conserved XPGG motif in the methyltransferase region is colored green and marked with ▲symbol. Secondary structure elements of N and C terminal domain are colored blue and red respectively for nsP2CHIKV. Similar residues are colored labeled and conserved identical residues are colored red.

The active site cavity in all the three structures was found to be hydrophobic with basic charged amino acids (Figure 5.12). However in surface view of the structure the cavity was observed to be open (Figure 5.12 A-B) and exposed except that of in nsP2CHIKV (Figure 5.12 C). Ribbon diagram shows that the helix on which catalytic cysteine is present don't show much of the variation but in the β-strand region, which brings the catalytic His and Trp closer to the catalytic Cys there is much variation (Figure 5.13). The region is variable with extension of the loop in outward position in case of nsP2SINV. And this variation runs down to the α6 where in case of SINV (r.m.s.d 0.952) as well as VEEV (r.m.s.d 3.282) considerable change with respect to CHIKV is observed (Figure 5.14 A, B). The nsP2CHIKV structure has large number of water molecules near the active site cavity and the residues N76 and L200 might be developing hydrophobic bonds with these water molecules rendering the closed active site. In case of VEEV, there are few water molecules and in SINV no water molecules were observed, this could be the cause of closed active site in these two proteases.



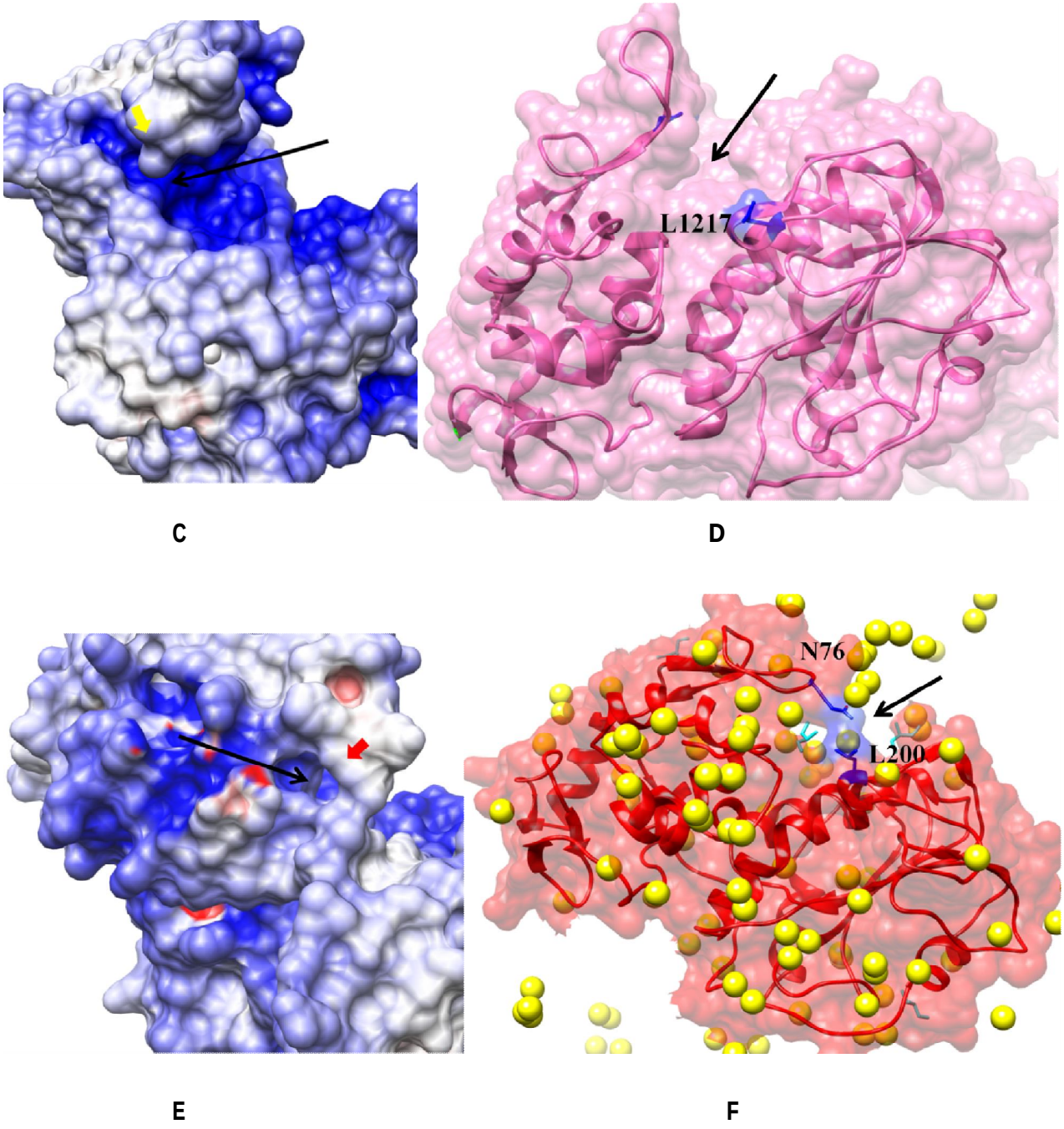


Figure 5.12 Electrostatic potential and surface view for VEEV, SINV and CHIKV nsP2. The view describes the cavity where substrate binds is hydrophobic in case of all the three members. Interestingly the cavity is open in case of VEEV (A,B) and SINV (C,D) whereas it appears to be closed in case of CHIKV (E, F). The cavity is marked by the black arrow. Yellow arrows in (A) and (B) points towards the open active site and red arrow in (C) points towards the

closed cavity in case of nsP2CHIKV. A,C, E are based on the electrostatic potential in case of VEEV, SINV and CHIKV respectively. The yellow balls in case of B and F are water molecules.

N-terminal region of nsP2CHIKV has considerable variations from the N-terminus of VEEV (r.m.s.d 5.510) and SINV (r.m.s.d 9.694), especially region spanning from β2-strand to β5-strand are with the maximum variation. The active site region which includes α1, β3 strands is more variable in case of SINV (r.m.s.d 3.132) than VEEV (r.m.s.d 0.510). This could be because of less sequence similarity in this region of SINV with CHIKV and VEEV nsP2 due to the insertion of seven extra amino acids. The α6, α7, β4, β5 region which lies just opposite to the active site is also found to be showing more variations. In case of SINV the r.m.s.d observed in these regions is less 0.9832, in comparison to the VEEV (r.m.s.d 1.465). This could be because of less sequence conservation in this region. It could be a possibility that binding of substrate in the active site is affected by the overall configuration of this whole region and thus it need to be flexible. B-factor is observed high in case of these regions and this also indicates that this region is flexibility. B-factor or Debye-waller factor is a term used for protein structures which indicates the relative vibrational motions of different parts of protein structure. Now how this attribute of flexibility is helpful in the activity is still not clear.

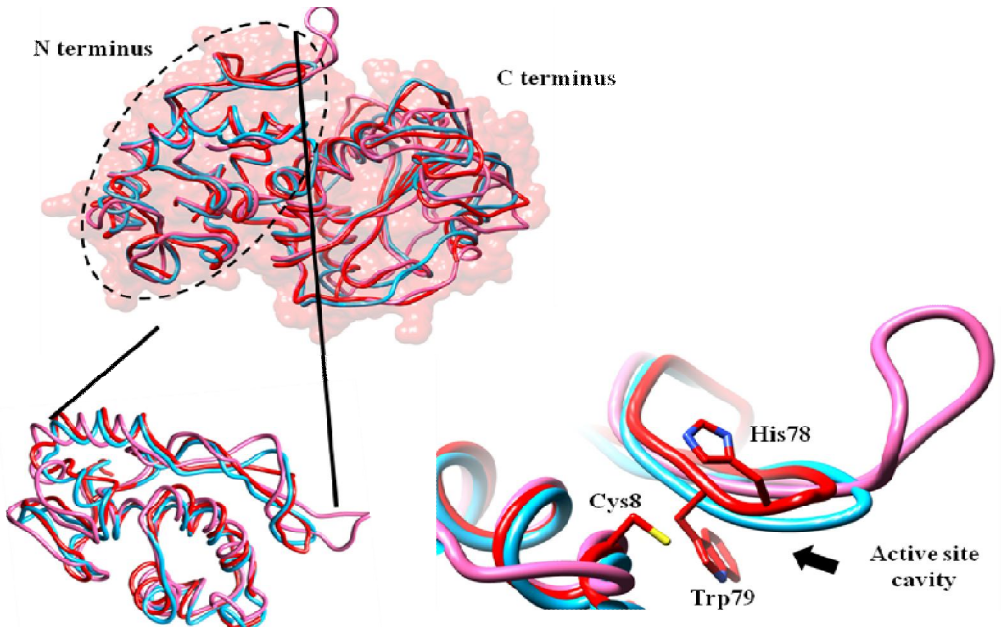


Figure 5.13 Comparative analysis of active site region of VEEV, SINV and CHIKV nsP2. The β-strand region on which the catalytic His and Trp are present is observed to be the most

flexible region around the active site. The region is marked by arrow here. VEEV (Blue), SINV (Pink) and CHIKV (Red).

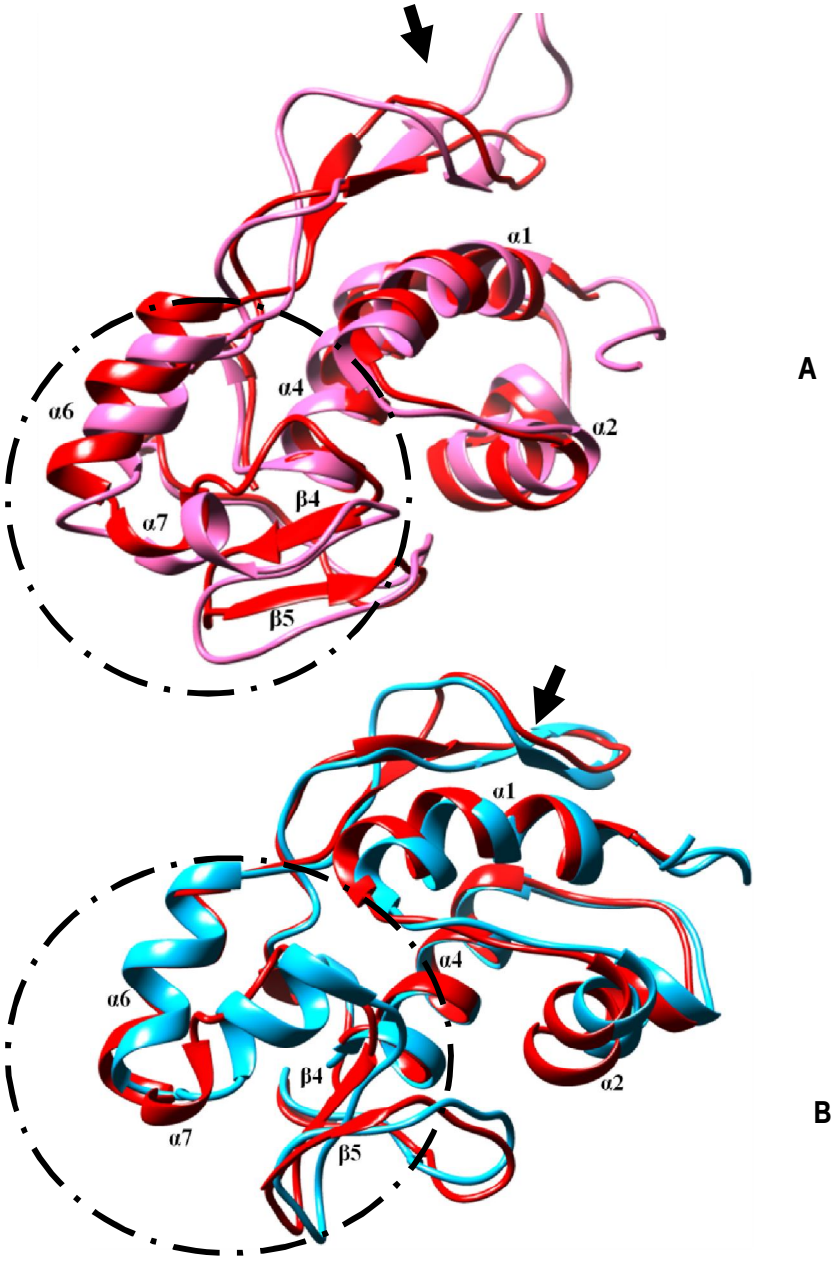


Figure 5.14 Superposition of N-terminal protease region. (A) Protease region of SINV (Pink) is superimposed with protease region of CHIKV (Red). **(B)** Protease region of VEEV (Blue) with CHIKV (Red). Region spanning from α6 to β5 is observed to be the most variable in both the alignments. R.m.s.d observed although still is more in case of SINV in this region in comparison to VEEV with respect to CHIKV (marked in green circle). Active site variable region is marked by arrow.

While comparing C-terminal region of the proteases overall conformation was observed to be conserved. The β-sheet region was completely overlapping except near α11 regions. α11 structure is overlapping in case of VEEV and CHIKV (r.m.s.d 0.810) but in case of SINV (r.m.s.d 1.310) this helix undergoes a rotation of around 8 degrees (Figure 5.15). Also the variation in β9, β6, β7 and helix α8 region is observed in case of SINV in comparison with CHIKV, but this region is not much variable when compared with VEEV (Figure 5.16 A, B).

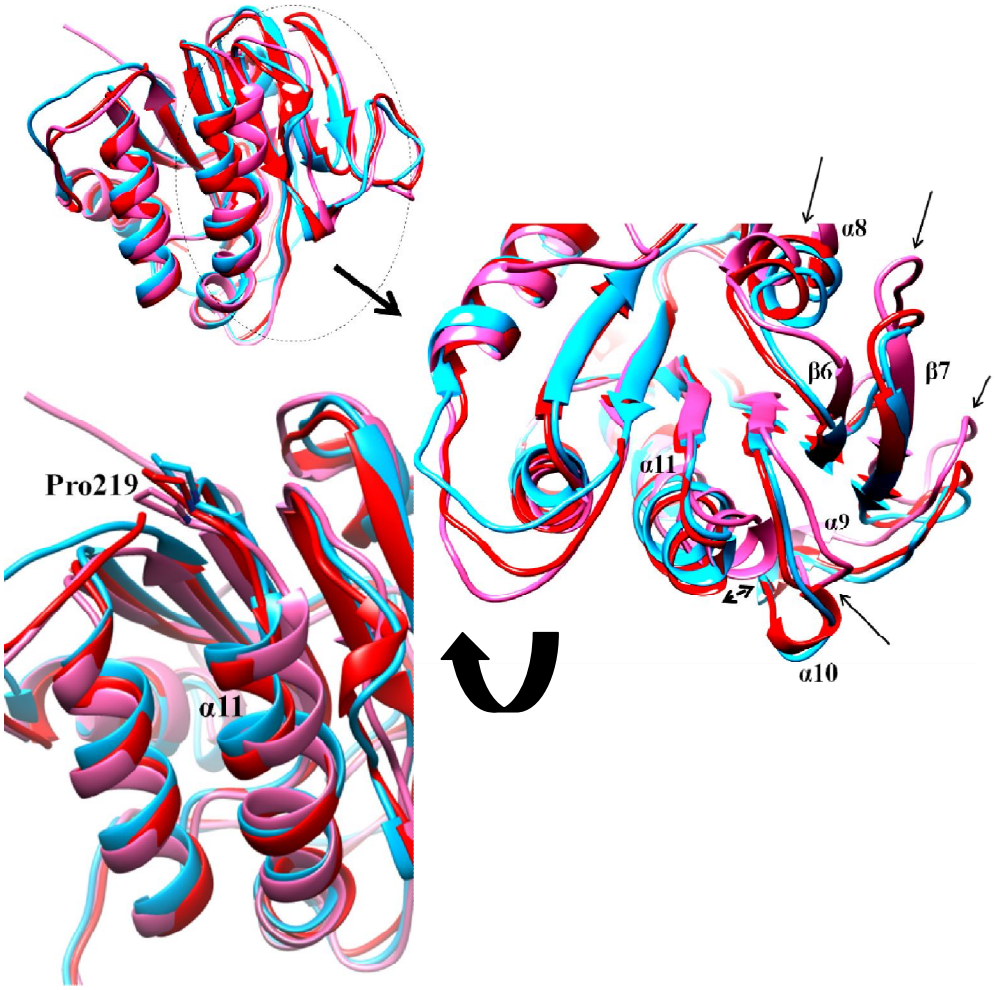


Figure 5.15 Comparison of methyltransferase region of VEEV, SINV and CHIKV. The α11 helix undergoes around 8 degree change in case of SINV w.r.t VEEV and CHIKV. The region spanning α11, α9, β6 and β7 in case of SINV is more variable than VEEV and CHIKV. The variation in α11 is marked by ↔ symbol. The Pro219 assumed to be interacting with the nsP3 is shown.

Since SINV structure was present with the nsP3ZBD region it could be possible that this region is variable and the interaction with the nsP3 might have caused this shift. But this shift affirms the flexibility and thus importance of this region.

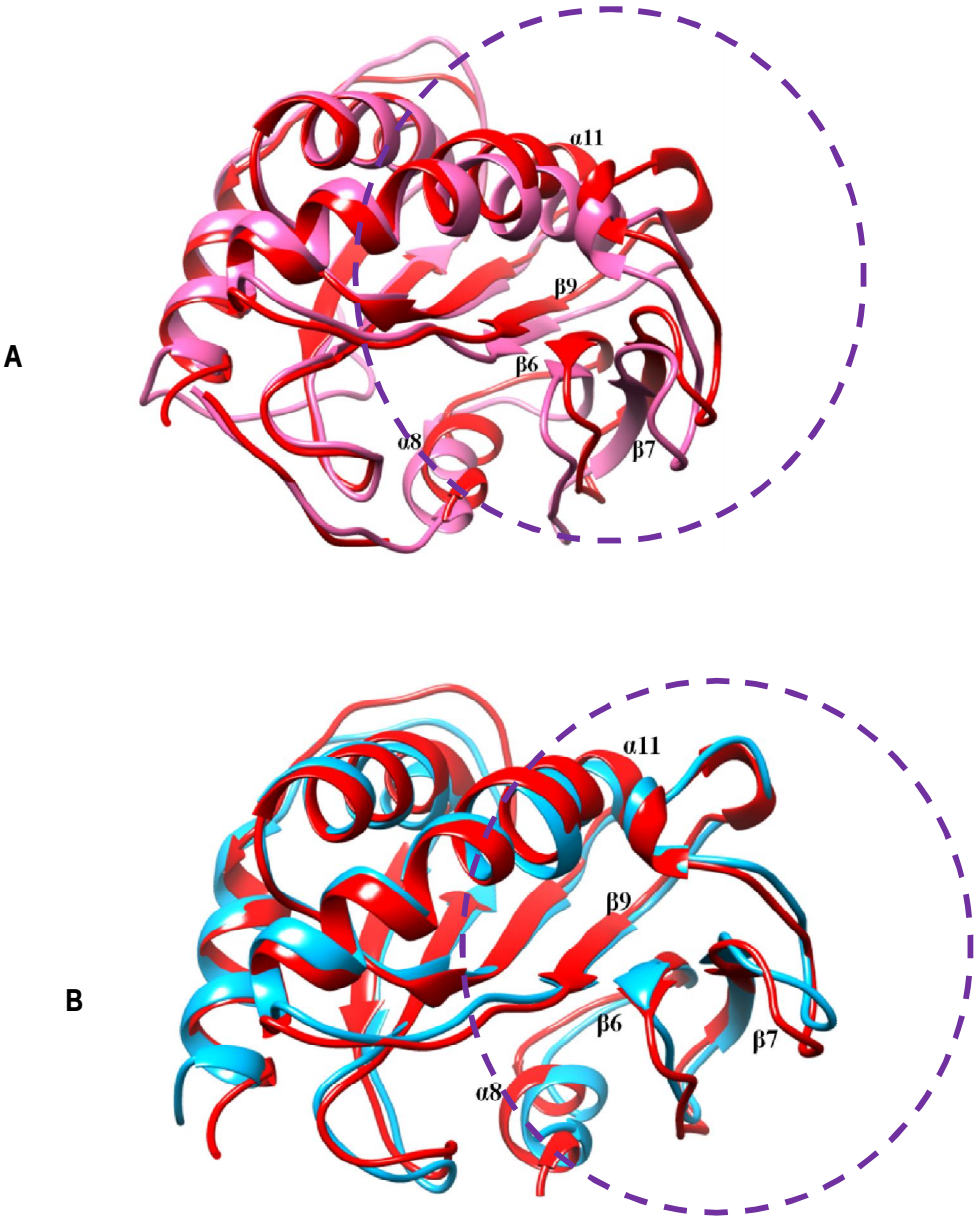


Figure 5.16 Alignment of C terminal methyltransferase region. (A) of SINV (Pink) with CHIKV (Red). (B) and of VEEV (Blue) with CHIKV (Red). The β sheet regions aligns perfectly in most of the places in case of both VEEV and SINV w.r.t CHIKV. However variations are observed as described in Figure 5.15. This figure clearly describes that SINV is more variable than VEEV from CHIKV and variable region are marked with dotted circle.

Pro258 present on the beta bulge connecting the α11 with β9 strand is conserved in all of the alphavirus members. This residue in case of SINV structure was found to be interacting with V162 and L165 of the nsP3 domain through a connecting loop. And any mutation of this Pro, negatively affects the sub-genomic and genomic RNA synthesis. This Pro is part of XPGG motif generally observed in case of RNA methyltransferases including FtsJ [7]. These variations alongwith these studies in the C terminal signifies the role of this domain in the RNA binding property of nsP2.

Another variation observed in these structures is observed in the active site region where the catalytic cysteine and histidine are observed to behave differently. In cysteine proteases, the Cys and His combine to form thiolinium-imidazolium ion pair during catalysis. Orientation of His imide ring is in plain with the Cys during the active state of the enzyme and since the ring can rotate up to 30 degree the ring might be able to undergo changes during different stages of the activity. However this rotation in papain proteases is seen to be controlled by a conserved Asn residue. But alphavirus cysteine proteases are devoid of any such third member of the catalytic triad. Instead it was observed that a conserved Trp is deemed important for the catalysis in alphaviral proteases [18].

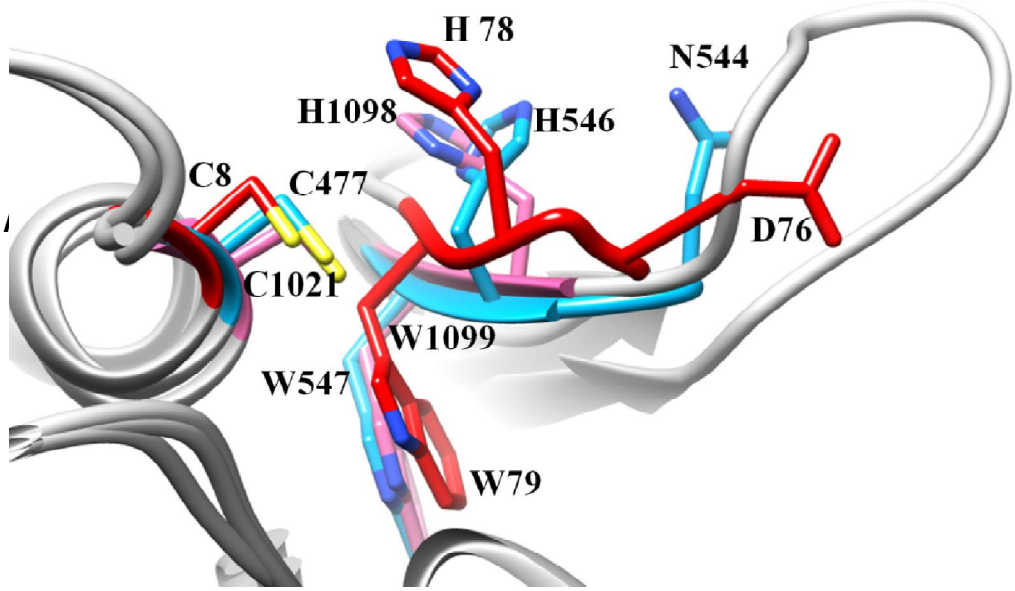
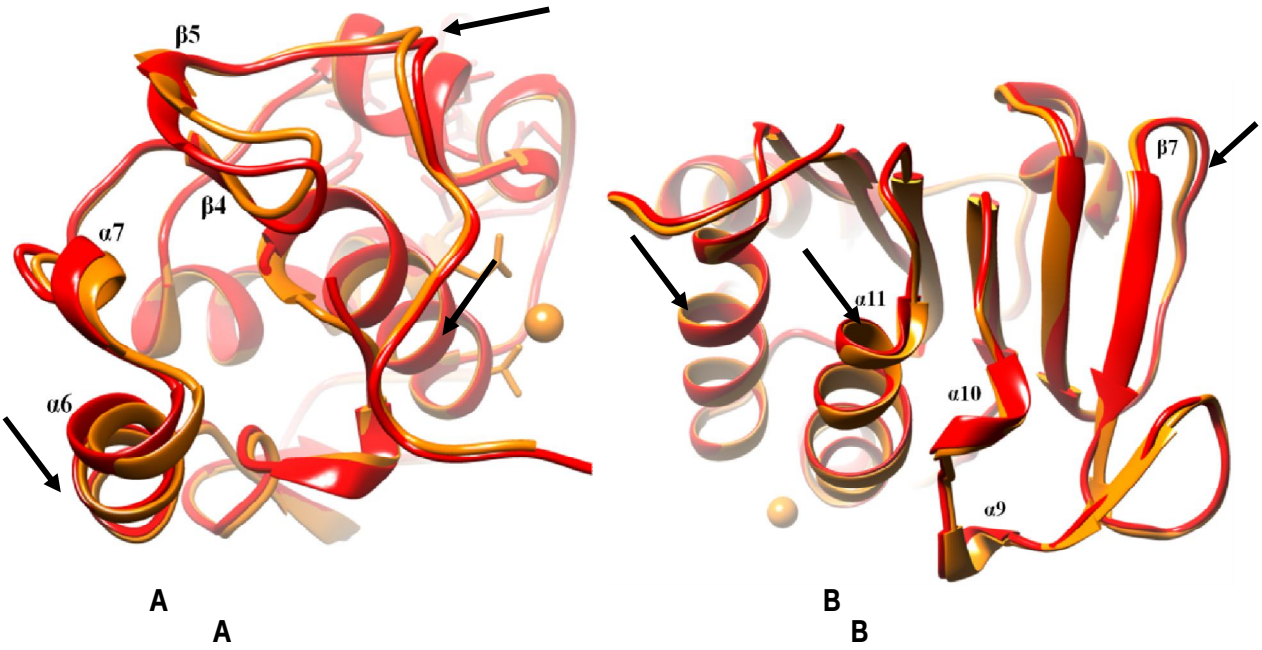


Figure 5.17. Orientation of catalytic Cys and His in all the three VEEV, SINV and CHIKV nsP2 as observed in the available X-ray structures. **(A)** In SINV (H1098) and CHIKV (H78) the imide ring is in plane with the catalytic Cys whereas it is rotated away in case of VEEV

(H546). His546 in case of nsP2VEEV is observed to be oriented towards N544 a conserved residue in alphaviruses.

When comparison of the structures of the available alphavirus non-structural proteases was done, it was observed that in CHIKV and SINV nsp2 structure the His ring is in plane with the Cys whereas it is rotated and moved outward in case of VEEV nsP2 structure (Figure 5.17 A). Also the His in case of VEEV is observed to be inclined towards Asn 544(Figure 5.17 B). The position of the His with respect to Cys in case of SINV and CHIKV might be depicting the active but resting state of the protein. In case of VEEV the imide ring of His might be oriented because of the presence of citrate in the active site cavity.

Comparing the 3TRK structure with nsP2CHIKV structure we have again observed variations. Surprisingly variations are observed in the regions where we have observed when compared with 2HWK and 4GUA structure. These variations point out that even when conserved sequentially there are certain regions which are flexible in the alphavirus non-structural proteases (Figure 5.18). This attribute of flexibility nsP2 is using for interaction with other proteins or how this is assisting in the activity of the protein is still to be deciphered. The active site region consisting of β3 strand shows r.m.s.d 0.987, however not considerable variation was observed in the α1 helix region. The N-terminal region consisting of the active site and catalytic residues have overall r.m.s.d 5.881, whereas the C terminal region shows r.m.s.d of only 0.772.



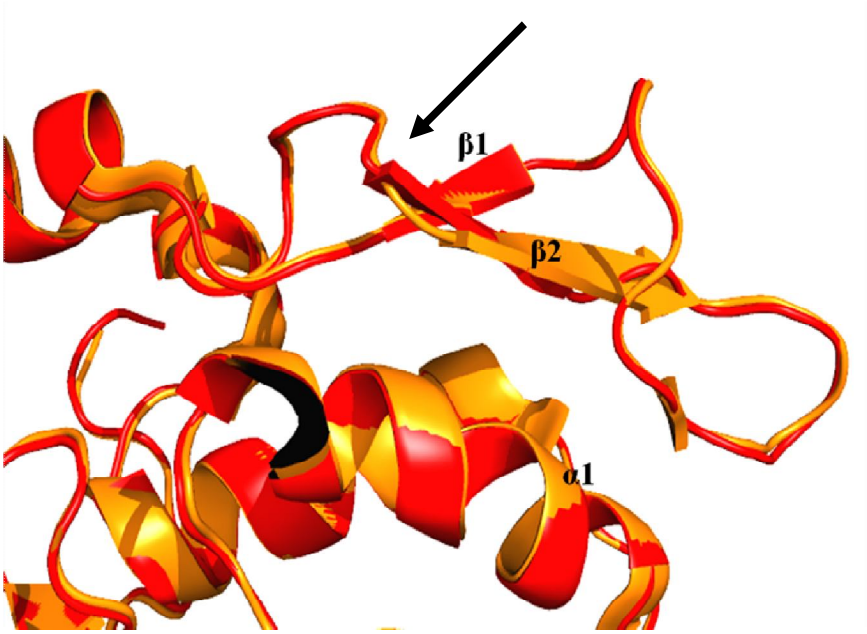


Figure 5.18 Comparison of nsP2CHIKV structure with 3TRK structure. Though being similar in sequence there are variation in the 3-D structure. The three regions where major changes are observed are shown here **(A)** the N terminal region **(B)** the methyltransferase region. and **(C)** the active site region. nsP2CHIKV (red), 3TRK (orange). Interestingly these are the major regions where variations are observed even in case of VEEV and SINV. Variable regions are marked by arrows.

5.4 Conclusion

The crystal structure for C-terminal protease domain of nsP2 from Chikungunya virus was solved at 2.56 Å. In solution the protein has always been observed to be monomer but here it was solved as a crystallographic tetramer. Description of the monomeric chain of nsP2CHIKV describing the secondary structure elements, types of bonds and unique features of this protein was given. Comparative analysis of different domains of this protein was done with the available structures of VEEV and SINV proteases. Significant variations in each domain with their possible implications were discussed. Also unique sequence features were also highlighted. Overall it was understood that there are regions in this protein which are dynamic and might be continuously undergoing certain changes to contribute in the different functions of the protein. However how these regions are helpful and how different residues present in these regions work is still to be deciphered. But mutational studies and understanding the dynamics of these regions might tell us more about this protein.

CHIKV is an important and pathogenic member of the alphavirus genus. Without any therapy and treatment it is a potent bioweapon and increasing demography of this virus confirms towards this possibility. Till now no complex structure of nsP2 protease from alphavirus is known and lack of any structural information of this type might hinder in the structure based drug discovery of inhibitors. Structure of nsP2CHIKV described here shows some of the important regions of the protein to be highly flexible. This information would contribute to the in silico method of developing the drug molecules. And would probably contribute towards the discovery of anti-nsP2 protease molecules which would help in treating CHIKV infections.

5.5 Bibliography

1. Suopanki, Jaana, Dorothea L. Sawicki, and Stanley G. Sawicki. "Regulation of alphavirus 26S mRNA transcription by replicase component nsP2." Journal of general virology 79, no. 2 (1998): 309-319.
2. Mayuri, Geders, Todd W., Janet L. Smith, and Richard J. Kuhn. "Role for conserved residues of Sindbis virus nonstructural protein 2 methyltransferase-like domain in regulation of minus-strand synthesis and development of cytopathic infection." Journal of virology 82, no. 15 (2008): 7284-7297.
3. Fros, Jelke J., Erika van der Maten, Just M. Vlak, and Gorben P. Pijlman. "The C-terminal domain of chikungunya virus nsP2 independently governs viral RNA replication, cytopathicity, and inhibition of interferon signaling." Journal of virology 87, no. 18 (2013): 10394-10400.
4. Shin, G., Yost, S. A., Miller, M. T., Elrod, E. J., Grakoui, A., & Marcotrigiano, J.. "Structural and functional insights into alphavirus polyprotein processing and pathogenesis." Proceedings of the National Academy of Sciences 109, no. 41 (2012): 16534-16539.
5. Barrett, Alan J., J. Fred Woessner, and Neil D. Rawlings, eds. Handbook of proteolytic enzymes. Vol. 1. Elsevier, 2012.
6. Kamphuis, Ireneus Gerhardus, K. H. Kalk, M. B. A. Swarte, and J_ Drenth. "Structure of papain refined at 1.65 Å resolution." Journal of molecular biology179, no. 2 (1984): 233-256.
7. Russo, Andrew T., Mark A. White, and Stanley J. Watowich. "The crystal structure of the Venezuelan equine encephalitis alphavirus nsP2 protease."Structure 14, no. 9 (2006): 1449-1458.
8. Russo, Andrew T., Robert D. Malmstrom, Mark A. White, and Stanley J. Watowich. "Structural basis for substrate specificity of alphavirus nsP2 proteases." Journal of Molecular Graphics and Modelling 29, no. 1 (2010): 46-53.
9. Polgar L. Mercaptide-imidazolium ion-pair: the reactive nucleophile in papain catalysis. FEBS Lett.1974;47:15–18.
10. Lulla A, Lulla V, Tints K, Ahola T, Merits A. Molecular determinants of substrate specificity for Semliki Forest virus nonstructural protease. J Virol. 2006;80:5413–5422.
11. Vasiljeva L, Merits A, Golubtsov A, Sizemskaja V, Kaariainen L, Ahola T. Regulation of the sequential processing of Semliki Forest virus replicase polyprotein. J Biol Chem. 2003;278:41636–41645.

12. Vasiljeva L, Valmu L, Kaariainen L, Merits A. Site-specific protease activity of the carboxyl-terminal domain of Semliki Forest virus replicase protein nsP2. J Biol Chem. 2001;276:30786–30793.
13. Shin, Gyehwa, Samantha A. Yost, Matthew T. Miller, Elizabeth J. Elrod, Arash Grakoui, and Joseph Marcotrigiano. "Structural and functional insights into alphavirus polyprotein processing and pathogenesis." Proceedings of the National Academy of Sciences 109, no. 41 (2012): 16534-16539.
14. Malet, Hélène, Bruno Coutard, Saïd Jamal, Hélène Dutartre, Nicolas Papageorgiou, Maarit Neuvonen, Tero Ahola et al. "The crystal structures of Chikungunya and Venezuelan equine encephalitis virus nsP3 macro domains define a conserved adenosine binding pocket." Journal of virology 83, no. 13 (2009): 6534-6545.
15. Rossmann, Michael G and David M Blow. "The Detection of Sub-Units within the Crystallographic Asymmetric Unit." Acta Crystallographica 15, no. 1 (1962): 24-31.
16. Jeffrey, George A., and George A. Jeffrey. An introduction to hydrogen bonding. Vol. 12. New York: Oxford university press, 1997.
17. Laskowski, Roman A. "PDBsum: summaries and analyses of PDB structures."Nucleic Acids Research 29, no. 1 (2001): 221-222.
18. Golubtsov A, Kaariainen L, Caldentey J. Characterization of the cysteine protease domain of Semliki Forest virus replicase protein nsP2 by in vitro mutagenesis. FEBS Lett. 2006;580:1502–1508.
19. Han, Y.S., Chang, G.G., Juo, C.G., Lee, H.J., Yeh, S.H., Hsu, J.T. and Chen, X. (2005) Papain-Like Protease 2 (PLP2) from Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV): Expression, Purification, Characterization, and Inhibition. Biochemistry 44:10349-10359.
20. Kanjanahaluethai, A. and Baker, S.C. (2000) Identification of mouse hepatitis virus papain-like proteinase 2 activity. J. Virol. 74:7911-7921.
21. Andrés, G., Alejo, A., Simón-Mateo, C. and Salas. M.L. (2001) African swine fever virus protease, a new viral member of the SUMO-1-specific protease family. J. Biol. Chem. 276:780-787.
22. Strauss, J. H., and E. G. Strauss. 1994. The alphaviruses: gene expression, replication, and evolution. Microbiol. Rev. 58:491-562. (Erratum, 58: 806.)

23. Sawicki, Dorothea L., Silvia Perri, John M. Polo, and Stanley G. Sawicki. "Role for nsP2 proteins in the cessation of alphavirus minus-strand synthesis by host cells." Journal of virology 80, no. 1 (2006): 360-371.
24. Griffin DE. Alphaviruses. In: Knipe DM, editor. Fields Virology. Vol 1. Philadelphia: Lippincott Williams & Wilkins; 2007. pp. 1023–1067.
25. Vasiljeva L, et al. Regulation of the sequential processing of Semliki Forest virus replicase polyprotein. J Biol Chem. 2003;278:41636–41645.
26. Vagin, Alexei and Alexei Teplyakov. "Molrep: An Automated Program for Molecular Replacement." Journal of Applied Crystallography 30, no. 6 (1997): 1022-1025.
27. Otwinowski, Z and W Minor. "Processing of X-Ray Diffraction Data." Methods enzymol 276, (1997): 307-326.
28. Murshudov, Garib N, Alexei A Vagin and Eleanor J Dodson. "Refinement of Macromolecular Structures by the Maximum-Likelihood Method." Acta Crystallographica Section D:Biological Crystallography 53, no. 3 (1997): 240-255.
29. Adams, Paul D, Pavel V Afonine, Gábor Bunkoczi, Vincent B Chen, Ian W Davis, Nathaniel Echols, Jeffrey J Headd, L-W Hung, Gary J Kapral and Ralf W Grosse-Kunstleve. "Phenix: A Comprehensive Python-Based System for Macromolecular Structure Solution." Acta Crystallographica Section D: Biological Crystallography 66, no. 2 (2010): 213-221.
30. Emsley, Paul and Kevin Cowtan. "Coot: Model-Building Tools for Molecular Graphics." Acta Crystallographica Section D: Biological Crystallography 60, no. 12 (2004): 2126-2132.