

STRUCTURAL AND FUNCTIONAL STUDIES OF CHIKUNGUNYA VIRUS CAPSID PROTEIN

Ph.D. THESIS

by

RAJESH SHARMA



**DEPARTMENT OF BIOTECHNOLOGY
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE – 247 667 (INDIA)
AUGUST, 2017**

STRUCTURAL AND FUNCTIONAL STUDIES OF CHIKUNGUNYA VIRUS CAPSID PROTEIN

A THESIS

*Submitted in partial fulfilment of the
requirements for the award of the degree*

of

DOCTOR OF PHILOSOPHY

in

BIOTECHNOLOGY

by

RAJESH SHARMA



**DEPARTMENT OF BIOTECHNOLOGY
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE – 247 667 (INDIA)
AUGUST, 2017**



**©INDIAN INSTITUTE OF TECHNOLOGY ROORKEE, ROORKEE-2017
ALL RIGHTS RESERVED**

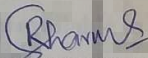


INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE

CANDIDATE'S DECLARATION

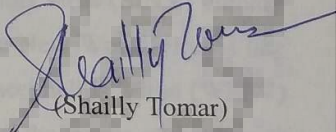
I hereby certify that the work which is being presented in this thesis entitled, "STRUCTURAL AND FUNCTIONAL STUDIES OF CHIKUNGUNYA VIRUS CAPSID PROTEIN" in partial fulfilment of the requirements for the award of the Degree of Doctor of Philosophy and submitted in the Department of Biotechnology of the Indian Institute of Technology Roorkee, Roorkee is an authentic record of my own work carried out during a period from July, 2012 to August, 2017 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 institution.

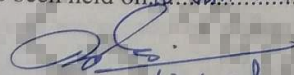

(RAJESH SHARMA)

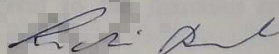
This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Dated: 10th Jan 2018

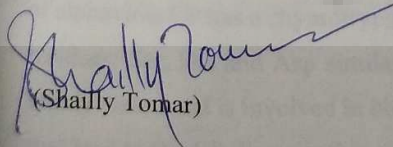

(Shailly Tomar)
Supervisor

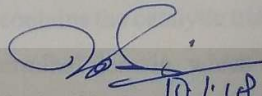
The Ph.D. Viva-voce Examination of Mr. Rajesh Sharma, Research Scholar, has been held on 10th Jan 2018.


Chairman, SRC 10.1.18


External Examiner

This is to certify that the student has made all the corrections in the thesis.


(Shailly Tomar)
Supervisor


Head of the Department 10.1.18

Abstract

Alphaviruses are a group of single-stranded positive-sense RNA viruses that are transmitted by mosquitoes and infect vertebrates. Alphaviruses belong to the *Togaviridae* family and include many medically relevant viruses like Chikungunya virus (CHIKV), Ross River virus (RRV), Western Equine Encephalitis virus (WEEV), Venezuelan Equine Encephalitis virus (VEEV) etc. Alphavirus infections cause fever, rash arthritis and encephalitis; and continues to be a potential threat to humans. The alphavirus virion is approximately 70 nm in diameter and has host cell derived lipid membrane which is embedded with 80 spikes in T=4 icosahedral symmetry. Each spike is composed of trimers of E1 and E2 glycoprotein heterodimers. The positive strand RNA genome (~11.7 kb) is surrounded by 240 copies of the capsid protein that forms a nucleocapsid core. The genomic RNA of virion has a 5' methylated cap and a 3' polyadenylated tail similar to host mRNA and on entry into the cytoplasm of host cell, viral genomic RNA is directly translated to viral proteins. The interaction of capsid protein (CP) with the cytoplasmic endodomain of E2 glycoprotein (cdE2) facilitate the budding of virus particles from the plasma membrane of host cell. Till now no effective antivirals have been developed against any member of Alphavirus genus nor a vaccine is commercially available in the market

Alphaviruses are classified into the New and the Old World viruses depending upon the mechanism by which they down regulate the transcription of host cell. The Old World viruses like Chikungunya virus (CHIKV), Semliki Forest virus (SFV), Sindbis virus (SINV) employ nsp2 protein for down regulating the cellular transcription, while the New World viruses like Venezuelan Equine Encephalitis Virus (VEEV), Western Equine Encephalitis Virus (WEEV) use capsid protein for shutting off the host transcription. Currently, the genus alphavirus is classified into seven antigenic complexes which includes both the New and Old World viruses. The La Reunion outbreak of chikungunya in 2005 infected about 40 % of the 785,000 population. This re-emergence of chikungunya infection poses a serious threat to the mankind and necessitates the study of alphaviral biology.

The CP of alphaviruses is a multifunctional protein. It consists of two domains, the RNA binding N-terminal domain and the C-terminal protease domain. The C-terminal protease domain of alphavirus CP has a chymotrypsin-like serine protease scaffold that contains the catalytic triad residues: Ser, His and Asp similar to other serine proteases. The N-terminal domain, which is highly disordered is involved in binding to the genomic RNA, PPIs (protein-protein interactions) that lead to the CP dimerization and the inhibition of host cell transcription. The C-terminal CP has *cis*-autoproteolytic activity and it cleaves itself from the N-terminus of the structural polyprotein. Thus, it plays a critical role in initiation of the structural polyprotein processing and

the viral life cycle. After the *cis*-cleavage, the conserved C-terminal tryptophan residue of CP remains bound to the active site and blocks further CP protease activity.

The molecular interaction of CP with the cdE2 facilitates the budding of virions from the plasma membrane of infected host cell. Previous studies have postulated that dioxane present in the CP-hydrophobic pocket of AVCP-dioxane complex structurally mimics the pyrrolidine ring of Pro405 residue of cdE2. This suggested that heterocyclic ring compounds similar to dioxane/compounds containing the pyrrolidine ring/ their derivatives can potentially bind in the CP hydrophobic pocket and disrupt its PPIs with cdE2. Picolinic acid (PCA), a pyridine containing compound is known to have antiviral, antimicrobial, cytotoxic and apoptotic properties. The binding of PCA to chikungunya virus capsid protein (CVCP) hydrophobic pocket was analysed by molecular docking, isothermal titration calorimetry (ITC), surface plasmon resonance (SPR) and fluorescence studies. Moreover, PCA significantly inhibited CHIKV replication in infected Vero cells, decreasing viral mRNA and viral load as assessed by qRT-PCR and plaque reduction assay, respectively.

The thesis consists of four chapters which include the structural and functional characterization of chikungunya virus capsid protein (CVCP). The structural and functional studies performed in this study have characterized the CVCP at the molecular level. These studies include *in silico* protein modeling, molecular docking, screening of small heterocyclic compounds targeting CP-cdE2 interaction, 3D crystal structure determination and analysis, CVCP *trans*-protease activity assay, beside biochemical and biophysical elucidations and interpretations.

Chapter 1 reviews the literature. It describes the genus alphavirus, life cycle, transmission, general structure of the virion. The genome organization of prototype alphavirus, structural and non-structural polyprotein processing have been described in details. The overall structure of the capsid protein and its multifunctional role in the virus life cycle like budding, autoproteolytic activity, nucleocapsid assembly and RNA encapsidation etc. have been described in details. Different strategies for CVCP protease inhibition and disruption of CP-cdE2 interactions have been discussed in the chapter.

Chapter 2 describes the *in silico* analysis of CVCP. Homology modeling was done for CVCP protein (residue range: 113-261). The crystal structure of Semliki Forest virus capsid protein (PDBID: 1VCP) was the first hit with 93 % sequence identity to CVCP. Dioxane based derivatives targeting protein-protein interactions have been reported to possess antiviral activity against Sindbis Virus (SINV), the prototype alphavirus. The modeled CVCP structure was used

for analyzing the binding of dioxane, proline and picolinic acid (PCA) into the conserved hydrophobic pocket of capsid protein. The recombinant CVCP has been produced in *E. coli* and purified using affinity chromatography and gel-filtration chromatography. The purified CVCP protein was used for biophysical studies. The binding of small heterocyclic compounds to purified protein was further analysed by ITC, SPR and fluorescence studies. The binding constant K_D obtained for PCA was 2.1×10^{-7} M. Additionally, PCA inhibits the CHIKV replication in Vero cells, decreasing viral mRNA and viral load as assessed by qRT-PCR and plaque reduction assay, respectively.

Chapter 3 describes the crystal structure determination and analysis of CVCP. The purified CVCP protein was used for crystallization and screened by using Hampton protein crystallization screens. The purified CVCP was crystallized and mountable crystals were obtained in 15 days. The addition of Octyl β -D-glucopyranoside to purified CVCP protein helped in improving the CVCP crystal quality. The crystals were diffracted at 2.0 Å, and data was collected at the home source (MCU, IIT Roorkee). The CVCP crystal data was solved using molecular replacement method and three dimensional structure analysis was done. The crystal structure reveals the presence of chymotrypsin-like structure having Greek key motif. A detailed analysis of CVCP protease active site and hydrophobic pocket has been done. Structural comparisons with crystal structure of other alphavirus capsid proteins have been done to identify residues of the conserved hydrophobic pocket and residues of the pocket that interacts with E1. Additionally, structural and sequential mapping of CP epitopes has been done. CVCP crystal structure was also used for structure-based identification of inhibitors targeting the conserved hydrophobic pocket of CVCP. Molecular docking of the compounds into the hydrophobic pocket of the CVCP crystal structure was performed. The binding affinity of these compounds to CVCP was further analyzed by surface plasmon resonance and fluorescence studies. Kinetic parameters of binding for these compounds to purified CVCP has been determined.

Chapter 4 reports the *trans*-proteolytic activity of alphavirus capsid protease. Virus specific proteases are verified drug targets including the example of classical HIV protease. However, for screening and testing of protease inhibitors we need a high throughput screening (HTS) protease assay. CHIKV capsid protease is an attractive target for anti-CHIKV agents because of its important role in structural polyprotein processing and nucleocapsid core formation inside host cells. This assay involves the use of fluorogenic peptide substrates for developing HTS assay. The sequence of the peptide was derived from the capsid protease cleavage site that includes the conserved W/S residues. When the capsid protease cleavage site

incorporated in the fluorogenic peptide substrate is cleaved by purified and active capsid protease, the FRET decreases and the fluorescence emission intensity for the donor fluorophore increases. Therefore, the protease activity of purified CP was detected by monitoring the increase of fluorescence at 490 nm. This fluorogenic peptide assay could be used for high throughput screening (HTS) of inhibitors against CHIKV capsid protease. Kinetic parameters using fluorogenic peptide substrates for the chikungunya virus capsid protease were estimated. The effect of serine protease inhibitors (N-p-Tosyl-L-phenylalanine chloromethyl ketone (TPCK) and Benzamidine hydrochloride) on CVCP protease activity has been determined. In conclusion, this chapter explains the development of a high-throughput method for testing and screening inhibitors against the proteolytic activity of CHIKV capsid protease.



ACKNOWLEDGEMENTS

I would like to convey my gratefulness to many people for my present thesis work. Foremost, I express my deep sense of gratitude to my supervisor, Dr. Shailly Tomar. Since I began my graduate studies, I am enlightened by her passion for science. Her painstaking guidance, unflagging interest, constructive criticism and perpetual encouragement at various stages shaped my work to be completed perfectly. Without her meticulous attention, the thesis work would not have been a success. Dr. Shailly Tomar provided me all the necessary facilities to ensure the successful voyage of my Ph. D. She has been my primary resource for getting my science questions answered and has been instrumental to crank out my thesis work. She has supported me academically and emotionally throughout my thesis. I could not have imagined having a better advisor and mentor for my Ph. D. study.

The valuable scientific studies are always the result of collaborative efforts. I endow my heartfelt thanks to Dr. Pravindra Kumar, for helping me out in the structural studies and giving me an authoritative support for the progress of my work. He is a polymath and the smartest person I know so far who encouraged me to do better. I couldn't have succeeded without his erudite support in structure biology and bioinformatics. I would like to thank him for his insightful discussions, suggestions and for all his scientific advices which provided me positive outlook during different phases of journey of my Ph. D.

I acknowledge with thanks the kind of patronage, loving inspiration and timely guidance which I received from my committee members Prof. R. P. Singh, Prof. Partha Roy and Dr. P. Gopinath. I am very fortunate to get the opportunity of having such great committee members.

Some faculty members have been very kind to extend their help at various stages of my research. I am grateful to Prof. G. S. Randhawa, Prof. Vikas Pruthi, Dr. A. K. Sharma, Dr. Naveen K. Navani, Dr. Ranjana Pathania, Dr. Harsh Chauhan, Dr. Kiran, Dr. Mandal, Dr. Deepak Sharma, Dr. Krishnanmohan Poluri, Dr. Sircar, Dr. Soma Rohtangi for their help and encouragement.

I owe my sincere thanks to Dr. Sonali Dhindwal, Dr. Megha Aggarwal, Dr. Satya Tapas, Dr. Preeti, Dr. Bibekananda Kar, Dr. Dipak Patil, Dr. Preeti Verma, Dr. Rajnikanth Sharma Dr. Manju Narwal, Dr. Monu Batra to be instrumental for helping me to get accustomed to the laboratory. I had a great time in laboratory with Dr. Sonali Dhindwal, Dr. Megha Aggarwal and Dr. Bibekananda Kar who helped me a lot as seniors and friends as well.

It is a humbling experience to acknowledge those who have, mostly out of kindness helped along the journey of my Ph. D. I thank, Ms. Pooja Kesari, Mr. Rajat Mudgal, Mr. Madhusudhan Rao Katiki, Mr. Harvijay, Ms. Ramanjit, Ms. Benazir Fatma for the fruitful discussions. I appreciate the discussions with Dr. Sonali Dhindwal, Ms. Pooja Kesari and Mr. Rajat Mudgal who helped me whenever I was not getting positive results during my research

work. I am thankful to all my labmates to facilitate the fun environment in the laboratory and also for helping me in all the possible way during my tenure.

I would like to thank Mr. Harvijay, Ms. Ramanjit, Mr. Pranav Kumar, Ms. Anjali Malik, Ms. Benazir Fatma, Mr. Akshay Pareek, Ms. Vedita Anand Singh, Ms. Gunjan Saini, Ms. Anchal Sharma, Mr. Vijay Sharma, Ms. Neha Singh, Ms. Neetu, Dr. Ravi Saini, Ms. Supreeti Mahajan and Ms. Happy Goyal for their sincere cooperation and efforts.

I must also thank Dr. Sakshi Sharma, Mr. Rajat Mudgal and Ms. Pooja Kesari for always being there with a word of encouragement or listening ear during my tough phases of Ph. D. This thesis would not exist without their love and support. I am also grateful to the Institute Sports Club as well as the Institute Hospital. Both facilitated me in maintaining physical fitness so that I could achieve my studies contained in this thesis.

Finally, and most importantly, my heartiest thank goes to my parents and all of my family members for their blessings and unconditional love. Their support and encouragement was worth more than I can express on a paper. They encouraged and supported me along the bumpy roads of my research work and inspire me to achieve the targets. They always stand firmly behind me and provide resources to carry out my studies. I would like to express my heartfelt gratitude to my family. None of this would have been possible without the love and the patience of my family. They have cherished with me every great moment and supported me whenever I needed it.

Beyond all these, I thank to almighty God for his blessings on me and keeping me strong and calm during all good and hard times throughout my life.

Finally I would like to thank to Department of Biotechnology (DBT), New Delhi India for financial assistance as JRF and SRF.

(Rajesh)

LIST OF PUBLICATIONS

1. Monu Batra, **Rajesh Sharma**, Vemika Chandra, Megha Aggarwal, Uday Agarwal, Pawan Gupta, Rajesh Pratap Singh, and Shailly Tomar. "In silico and proteomic analysis of protein

methyltransferase CheR from *Bacillus subtilis*." **International journal of biological macromolecules** 77 (2015): 168-180.

2. Megha Aggarwal, **Rajesh Sharma**, Pravindra Kumar, Manmohan Parida, Shailly Tomar. "Kinetic characterization of trans-proteolytic activity of Chikungunya virus capsid protease and development of a FRET-based HTS assay". **Scientific reports** 5 (2015).

3. Monu Batra¹, **Rajesh Sharma**¹, Anjali Malik, Sonali Dhindwal, Pravindra Kumar, and Shailly Tomar. "Crystal structure of pentapeptide-independent chemotaxis receptor methyltransferase (CheR) reveals idiosyncratic structural determinants for receptor recognition." **Journal of Structural Biology** 196, no. 3 (2016): 364-374.

4. **Rajesh Sharma**, Benazir Fatma, Amrita Saha, Sailesh Bajpai, Srinivas Sistla, Paban Kumar Dash, Manmohan Parida, Pravindra Kumar, and Shailly Tomar. "Inhibition of chikungunya virus by picolinate that targets viral capsid". **Virology** 498 (2016): 265-276.

5. **Rajesh Sharma**, Pooja Kesari, Pravindra Kumar and Shailly Tomar. "Structure-function insights into chikungunya virus capsid protease: Small molecules targeting capsid hydrophobic pocket". **Virology** 515 (2018): 223-234.

6. Benazir Fatma, **Rajesh Sharma**, Pooja Kesari, Shailly Tomar. "Structure based screening and evaluation of chikungunya virus capsid protease inhibitors for drug repurposing". (Under preparation).

CONTENTS

Page No.

CANDIDATE'S DECLARATION

ABSTRACT.....	i-iv
ACKNOWLEDGEMENTS.....	v-vi
LIST OF PUBLICATIONS.....	vii
CONTENTS.....	viii-xi
LIST OF FIGURES.....	xii-xiii
LIST OF TABLES.....	xiv
LIST OF ABBREVIATIONS USED.....	xv-xvi

CHAPTER 1

REVIEW OF LITERATURE

1.1	Introduction.....	1
1.2	Alphavirus life cycle.....	3
1.3	Structure of alphavirus virion.....	5
1.4	Genome architecture.....	6
1.5	Non-structural proteins.....	7
1.6	Structural proteins.....	10
1.7	Nucleocapsid assembly.....	14
1.8	Alphavirus budding.....	16
1.9	Role of glycoproteins in virus budding.....	17
1.10	Capsid protein structure.....	17
1.11	Alphavirus CP <i>trans</i> -activity.....	20

CHAPTER 2

INHIBITION OF CHIKUNGUNYA VIRUS BY PICOLINATE THAT TARGETS VIRAL CAPSID PROTEIN

2.1	Abstract.....	22
2.2	Introduction.....	23
2.3	Material and Methods.....	26

2.3.1	3D model generation.....	26
2.3.2	Molecular Docking.....	26
2.3.3	Virus.....	27
2.3.4	Expression and Purification of CVCP protein.....	27
	2.3.4.1 Expression of CVCP.....	28
	2.3.4.2 Purification of CVCP.....	28
2.3.5	Circular dichroism spectroscopy.....	29
2.3.6	Isothermal titration calorimetry.....	29
2.3.7	Surface plasmon resonance.....	30
	2.3.7.1 Buffer Preparation.....	30
	2.3.7.2 Immobilization of CVCP protein.....	30
2.3.8	Fluorescence spectroscopy.....	31
2.3.9	Cells and virus propagation.....	31
	2.3.9.1 Cytotoxicity testing.....	32
	2.3.9.2 Antiviral assay.....	32
	2.3.9.2.1 qRT-PCR.....	32
	2.3.9.2.2 Plaque reduction assay.....	33
2.4	Results.....	33
	2.4.1 3D Homology model.....	33
	2.4.2 Molecular docking.....	36
	2.4.3 Purification of CVCP and secondary structure analysis.....	38
	2.4.4 Isothermal studies of picolinic acid binding.....	38
	2.4.5 Surface plasmon resonance.....	39
	2.4.5.1 CVCP protein immobilization.....	40
	2.4.5.2 Kinetic analysis.....	40
	2.4.6 Effect of PCA on CVCP intrinsic fluorescence.....	41
	2.4.7 Cytotoxicity testing.....	42
	2.4.8 Antiviral assay.....	43
2.5	Discussion.....	45
2.6	Conclusions.....	48

CHAPTER 3

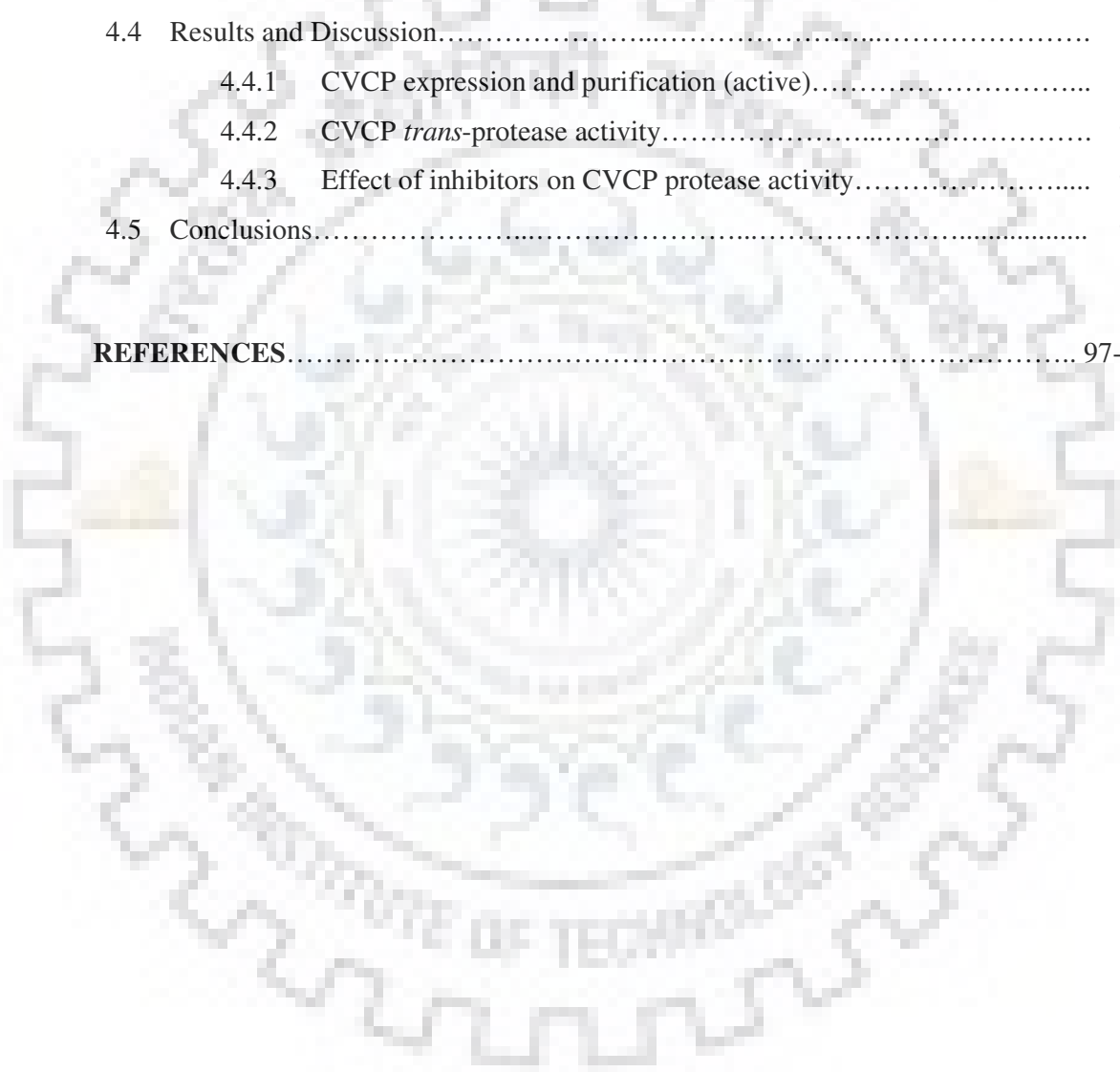
STRUCTURE-FUNCTION INSIGHTS INTO CHIKUNGUNYA VIRUS CAPSID PROTEIN: SMALL MOLECULES TARGETING CAPSID HYDROPHOBIC POCKET

3.1	Abstract.....	49
3.2	Introduction.....	50
3.3	Material and Methods.....	53
3.3.1	Expression and Purification.....	53
3.3.2	Crystallization and data collection.....	53
3.3.3	Structure solution and refinement.....	54
3.3.4	Molecular modeling.....	55
3.3.5	Identification and docking of small heterocyclic molecules.....	55
3.3.6	Surface plasmon resonance (SPR)	56
3.3.7	Fluorescence spectroscopy.....	56
3.3.8	<i>In-silico</i> epitope identification.....	57
3.4	Results and Discussion.....	57
3.4.1	Structure determination and overall architecture.....	57
3.4.2	Monomer architecture.....	59
3.4.3	Substrate binding site architecture.....	60
3.4.4	Dimer architecture.....	61
3.4.5	CVCP-cdE2 glycoprotein interactions.....	64
3.4.6	Structural comparison of CVCP with other alphavirus CP's.....	68
3.4.7	Identification and docking of small heterocyclic molecules.....	70
3.4.8	Surface plasmon resonance.....	73
3.4.9	Effect of MDA and EAB on CVCP intrinsic fluorescence.....	75
3.4.10	Epitope identification.....	76
3.5	Conclusion.....	78

CHAPTER 4

CHARACTERIZATION OF TRANS-PROTEOLYTIC ACTIVITY OF CHIKUNGUNYA VIRUS CAPSID PROTEASE

4.1	Abstract.....	80
4.2	Introduction.....	81
4.3	Material and Methods.....	86
4.3.1	Expression of active CVCP.....	86
4.3.2	Purification of active CVCP.....	87
4.3.3	CVCP <i>trans</i> -protease activity.....	87
4.3.4	Effect of inhibitors on CVCP protease activity.....	89
4.4	Results and Discussion.....	89
4.4.1	CVCP expression and purification (active).....	89
4.4.2	CVCP <i>trans</i> -protease activity.....	91
4.4.3	Effect of inhibitors on CVCP protease activity.....	93
4.5	Conclusions.....	95
	REFERENCES.....	97-120



LIST OF FIGURES

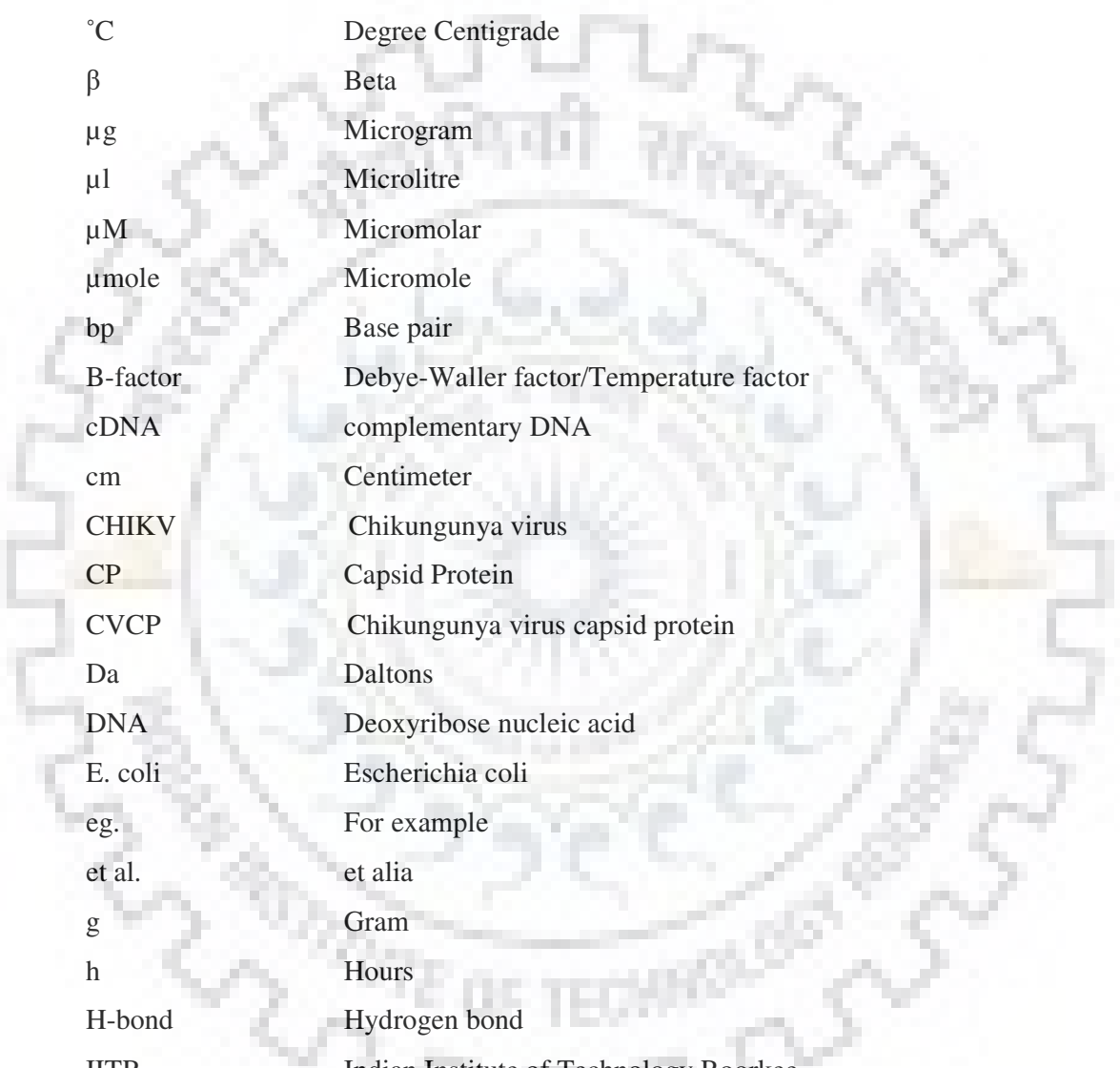
Figure 1.2.1:	Graphical illustration of the alphavirus transmission cycle.....	3
Figure 1.2.2:	Schematic representation of alphavirus life cycle.....	4
Figure 1.3.1:	Virion structure organization.....	5
Figure 1.4.1:	Genome organization.....	6
Figure 1.5.1:	Non-structural polyprotein processing.....	8
Figure 1.6.1:	Processing of structural polyprotein.....	11
Figure 1.6.2:	Different regions of alphavirus capsid protein.....	12
Figure 1.7.1:	Nucleocapsid assembly.....	14
Figure 1.8.1:	Schematic representation of two types of budding models.....	16
Figure 1.10.1:	Cartoon view representation of alphavirus crystal structures.....	19
Figure 2.4.1.1:	3D atomic model structure of CVCP.....	34
Figure 2.4.1.2:	Structure comparison with alphavirus capsid protein.....	34
Figure 2.4.1.3:	Multiple sequence alignment of CVCP protein with sequences of other alphavirus capsid protein (CP).....	36
Figure 2.4.2.1:	Binding analysis of docked compound into CVCP hydrophobic pocket.....	37
Figure 2.4.3.1:	SDS-PAGE analysis of chikungunya virus capsid protein (CVCP)...	38
Figure 2.4.4.1:	ITC raw and integrated data.	39
Figure 2.4.5.1:	SPR binding and kinetic analysis.....	41
Figure 2.4.6.1:	Fluorescence spectroscopy.....	42
Figure 2.4.7.1:	Apoptosis in PCA treated Vero cells.....	43
Figure 2.4.8.1:	qRT-PCR.....	44
Figure 2.4.8.2:	Plaque reduction assay.....	45
Figure 3.3.1.1:	Gel filtration chromatography and SDS-PAGE analysis of CVCP.....	53
Figure 3.3.2.1:	Crystal and Diffraction image of CVCP.....	54
Figure 3.4.2.1:	Crystal structure of CVCP.....	60
Figure 3.4.3.1:	Catalytic architecture of CVCP.....	61
Figure 3.4.4.1:	Dimer interface interactions.....	62
Figure 3.4.4.2:	Structural insights of dimer interface interactions.....	63

Figure 3.4.4.3:	Multiple sequence alignment of CVCP with CP's from other alphaviruses.....	64
Figure 3.4.5.1:	CVCP crystal structure and E1, E2 glycoprotein homology models fitted in to cryo-EM density map of VEEV.....	65
Figure 3.4.5.2:	Depiction of Helix-loop-Helix of cdE2 glycoprotein fitted in to CVCP hydrophobic pocket.....	66
Figure 3.4.5.3:	CVCP- cdE2 interactions.....	67
Figure 3.4.5.4:	Multiple sequence alignment of cytoplasmic domain of E2 glycoprotein (cdE2) from different alphaviruses.....	67
Figure 3.4.5.5:	Hydrophobic interactions of Pro404 of cdE2 with CVCP hydrophobic pocket.....	68
Figure 3.4.6.1:	Superimposition of CP's from different alphaviruses.....	70
Figure 3.4.7.1:	Binding analysis of docked (S)-(+)-Mandelic acid (MDA) into CVCP hydrophobic pocket.....	72
Figure 3.4.7.2:	Binding analysis of docked Ethyl 3-aminobenzoate (EAB) into CVCP hydrophobic pocket.....	72
Figure 3.4.8.1:	SPR binding and kinetic analysis of (S)-(+)-Mandelic acid.....	73
Figure 3.4.8.2:	SPR binding and kinetic analysis of Ethyl 3-aminobenzoate.....	74
Figure 3.4.9.1:	Intrinsic fluorescence spectroscopy analysis of MDA.....	75
Figure 3.4.9.2:	Intrinsic fluorescence spectroscopy analysis of MDA.....	75
Figure 3.4.10.1:	<i>In-silico</i> epitope identification.....	77
Figure 4.2.1:	Catalytic mechanism of serine proteases.....	84
Figure 4.2.2:	The LigPlot analysis of benzamidine binding to protease S1 pocket...	85
Figure 4.2.3:	Mechanism of inactivation of serine and cysteine proteases by halomethyl ketone.....	86
Figure 4.3.3.1:	Fluorescence resonance energy transfer (FRET) protease assay.....	88
Figure 4.4.1.1:	Schematic representation of alphavirus active and inactive protease...	90
Figure 4.4.1.2:	Expression of active CVCP.....	91
Figure 4.4.1.3:	Purification of active CVCP.....	91
Figure 4.4.2.1:	Kinetic analysis of active CVCP.....	92
Figure 4.4.3.1:	Determination of inhibition constant for BHC	94
Figure 4.4.3.2:	Determination of inhibition constant for TPCK.....	95


LIST OF TABLES

Table 1.1: Functions of non-structural proteins (nsPs).....	10
Table 1.2: Functions of structural proteins.....	13
Table 1.3: List of CP crystal structures from different members of alphavirus family	18
Table 2.1: List of the compounds with their empirical formula and molecular structure used for docking into the hydrophobic pocket of CVCP.....	27
Table 2.2: The various set parameters used for performing the Isothermal titration calorimetry experiments of dioxane and picolinic acid with purified CVCP.....	30
Table 2.3: The values of binding energy obtained from molecular docking of dioxane, proline and picolinic acid into the CVCP hydrophobic pocket....	37
Table 2.4: The thermodynamic analysis of dioxane and picolinic acid binding to CVCP as obtained from ITC.....	39
Table 3.1: Data collection and refinement statistics for CVCP.....	58
Table 3.2: CVCP intra-subdomain salt bridges.....	59
Table 3.3: Interface polar interactions between the two crystallographic monomers of CVCP.....	63
Table 3.4: The molecular docking results of dioxane, picolinic acid, (S)-(+)-Mandelic acid and Ethyl 3-aminobenzoate binding to CVCP hydrophobic pocket.....	71

LIST OF ABBREVIATIONS USED



Å	Angstrom
α	Alpha
°C	Degree Centigrade
β	Beta
μg	Microgram
μl	Microlitre
μM	Micromolar
μmole	Micromole
bp	Base pair
B-factor	Debye-Waller factor/Temperature factor
cDNA	complementary DNA
cm	Centimeter
CHIKV	Chikungunya virus
CP	Capsid Protein
CVCP	Chikungunya virus capsid protein
Da	Daltons
DNA	Deoxyribose nucleic acid
E. coli	Escherichia coli
eg.	For example
et al.	et alia
g	Gram
h	Hours
H-bond	Hydrogen bond
IITR	Indian Institute of Technology Roorkee
IPTG	Isopropyl β-D-thiogalactoside
ITC	Isothermal titration calorimetry
K	Kelvin
kDa	Kilo Daltons
m	Meter
M	Molar
mg	milligram



min	Minute
ml	millilitre
mm	millimeter
mM	millimolar
NaCl	Sodium Chloride
NCBI	National Center for Biotechnology Information
NI-NTA	Nickel-nitrilotriacetic acid
nm	Nanometer
PCR	Polymerase chain reaction
PDB	Protein Data Bank
R-factor	Residual-factor
RNA	Ribose nucleic acid
rpm	Revolutions per minute
s	Seconds
SPR	Surface plasmon resonance
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
v/v	volume /volume

Abstract

Alphaviruses are a group of single-stranded positive-sense RNA viruses that are transmitted by mosquitoes and infect vertebrates. Alphaviruses belong to the *Togaviridae* family and include many medically relevant viruses like Chikungunya virus (CHIKV), Ross River virus (RRV), Western Equine Encephalitis virus (WEEV), Venezuelan Equine Encephalitis virus (VEEV) etc. Alphavirus infections cause fever, rash arthritis and encephalitis; and continues to be a potential threat to humans. The alphavirus virion is approximately 70 nm in diameter and has host cell derived lipid membrane which is embedded with 80 spikes in T=4 icosahedral symmetry. Each spike is composed of trimers of E1 and E2 glycoprotein heterodimers. The positive strand RNA genome (~11.7 kb) is surrounded by 240 copies of the capsid protein that forms a nucleocapsid core. The genomic RNA of virion has a 5' methylated cap and a 3' polyadenylated tail similar to host mRNA and on entry into the cytoplasm of host cell, viral genomic RNA is directly translated to viral proteins. The interaction of capsid protein (CP) with the cytoplasmic endodomain of E2 glycoprotein (cdE2) facilitate the budding of virus particles from the plasma membrane of host cell. Till now no effective antivirals have been developed against any member of Alphavirus genus nor a vaccine is commercially available in the market

Alphaviruses are classified into the New and the Old World viruses depending upon the mechanism by which they down regulate the transcription of host cell. The Old World viruses like Chikungunya virus (CHIKV), Semliki Forest virus (SFV), Sindbis virus (SINV) employ nsp2 protein for down regulating the cellular transcription, while the New World viruses like Venezuelan Equine Encephalitis Virus (VEEV), Western Equine Encephalitis Virus (WEEV) use capsid protein for shutting off the host transcription. Currently, the genus alphavirus is classified into seven antigenic complexes which includes both the New and Old World viruses. The La Reunion outbreak of chikungunya in 2005 infected about 40 % of the 785,000 population. This re-emergence of chikungunya infection poses a serious threat to the mankind and necessitates the study of alphaviral biology.

The CP of alphaviruses is a multifunctional protein. It consists of two domains, the RNA binding N-terminal domain and the C-terminal protease domain. The C-terminal protease domain of alphavirus CP has a chymotrypsin-like serine protease scaffold that contains the catalytic triad residues: Ser, His and Asp similar to other serine proteases. The N-terminal domain, which is highly disordered is involved in binding to the genomic RNA, PPIs (protein-protein interactions) that lead to the CP dimerization and the inhibition of host cell transcription. The C-terminal CP has *cis*-autoproteolytic activity and it cleaves itself from the N-terminus of the structural polyprotein. Thus, it plays a critical role in initiation of the structural polyprotein processing and

the viral life cycle. After the *cis*-cleavage, the conserved C-terminal tryptophan residue of CP remains bound to the active site and blocks further CP protease activity.

The molecular interaction of CP with the cdE2 facilitates the budding of virions from the plasma membrane of infected host cell. Previous studies have postulated that dioxane present in the CP-hydrophobic pocket of AVCP-dioxane complex structurally mimics the pyrrolidine ring of Pro405 residue of cdE2. This suggested that heterocyclic ring compounds similar to dioxane/compounds containing the pyrrolidine ring/ their derivatives can potentially bind in the CP hydrophobic pocket and disrupt its PPIs with cdE2. Picolinic acid (PCA), a pyridine containing compound is known to have antiviral, antimicrobial, cytotoxic and apoptotic properties. The binding of PCA to chikungunya virus capsid protein (CVCP) hydrophobic pocket was analysed by molecular docking, isothermal titration calorimetry (ITC), surface plasmon resonance (SPR) and fluorescence studies. Moreover, PCA significantly inhibited CHIKV replication in infected Vero cells, decreasing viral mRNA and viral load as assessed by qRT-PCR and plaque reduction assay, respectively.

The thesis consists of four chapters which include the structural and functional characterization of chikungunya virus capsid protein (CVCP). The structural and functional studies performed in this study have characterized the CVCP at the molecular level. These studies include *in silico* protein modeling, molecular docking, screening of small heterocyclic compounds targeting CP-cdE2 interaction, 3D crystal structure determination and analysis, CVCP *trans*-protease activity assay, beside biochemical and biophysical elucidations and interpretations.

Chapter 1 reviews the literature. It describes the genus alphavirus, life cycle, transmission, general structure of the virion. The genome organization of prototype alphavirus, structural and non-structural polyprotein processing have been described in details. The overall structure of the capsid protein and its multifunctional role in the virus life cycle like budding, autoproteolytic activity, nucleocapsid assembly and RNA encapsidation etc. have been described in details. Different strategies for CVCP protease inhibition and disruption of CP-cdE2 interactions have been discussed in the chapter.

Chapter 2 describes the *in silico* analysis of CVCP. Homology modeling was done for CVCP protein (residue range: 113-261). The crystal structure of Semliki Forest virus capsid protein (PDBID: 1VCP) was the first hit with 93 % sequence identity to CVCP. Dioxane based derivatives targeting protein-protein interactions have been reported to possess antiviral activity

against Sindbis Virus (SINV), the prototype alphavirus. The modeled CVCP structure was used for analyzing the binding of dioxane, proline and picolinic acid (PCA) into the conserved hydrophobic pocket of capsid protein. The recombinant CVCP has been produced in *E. coli* and purified using affinity chromatography and gel-filtration chromatography. The purified CVCP protein was used for biophysical studies. The binding of small heterocyclic compounds to purified protein was further analysed by ITC, SPR and fluorescence studies. The binding constant K_D obtained for PCA was 2.1×10^{-7} M. Additionally, PCA inhibits the CHIKV replication in Vero cells, decreasing viral mRNA and viral load as assessed by qRT-PCR and plaque reduction assay, respectively.

Chapter 3 describes the crystal structure determination and analysis of CVCP. The purified CVCP protein was used for crystallization and screened by using Hampton protein crystallization screens. The purified CVCP was crystallized and mountable crystals were obtained in 15 days. The addition of Octyl β -D-glucopyranoside to purified CVCP protein helped in improving the CVCP crystal quality. The crystals were diffracted at 2.0 Å, and data was collected at the home source (MCU, IIT Roorkee). The CVCP crystal data was solved using molecular replacement method and three dimensional structure analysis was done. The crystal structure reveals the presence of chymotrypsin-like structure having Greek key motif. A detailed analysis of CVCP protease active site and hydrophobic pocket has been done. Structural comparisons with crystal structure of other alphavirus capsid proteins have been done to identify residues of the conserved hydrophobic pocket and residues of the pocket that interacts with E1. Additionally, structural and sequential mapping of CP epitopes has been done. CVCP crystal structure was also used for structure-based identification of inhibitors targeting the conserved hydrophobic pocket of CVCP. Molecular docking of the compounds into the hydrophobic pocket of the CVCP crystal structure was performed. The binding affinity of these compounds to CVCP was further analyzed by surface plasmon resonance and fluorescence studies. Kinetic parameters of binding for these compounds to purified CVCP has been determined.

Chapter 4 reports the *trans*-proteolytic activity of alphavirus capsid protease. Virus specific proteases are verified drug targets including the example of classical HIV protease. However, for screening and testing of protease inhibitors we need a high throughput screening (HTS) protease assay. CHIKV capsid protease is an attractive target for anti-CHIKV agents because of its important role in structural polyprotein processing and nucleocapsid core formation inside host cells. This assay involves the use of fluorogenic peptide substrates for

developing HTS assay. The sequence of the peptide was derived from the capsid protease cleavage site that includes the conserved W/S residues. When the capsid protease cleavage site incorporated in the fluorogenic peptide substrate is cleaved by purified and active capsid protease, the FRET decreases and the fluorescence emission intensity for the donor fluorophore increases. Therefore, the protease activity of purified CP was detected by monitoring the increase of fluorescence at 490 nm. This fluorogenic peptide assay could be used for high throughput screening (HTS) of inhibitors against CHIKV capsid protease. Kinetic parameters using fluorogenic peptide substrates for the chikungunya virus capsid protease were estimated. The effect of serine protease inhibitors (N-p-Tosyl-L-phenylalanine chloromethyl ketone (TPCK) and Benzamidine hydrochloride) on CVCP protease activity has been determined. In conclusion, this chapter explains the development of a high-throughput method for testing and screening inhibitors against the proteolytic activity of CHIKV capsid protease.



ACKNOWLEDGEMENTS

I would like to convey my gratefulness to many people for my present thesis work. Foremost, I express my deep sense of gratitude to my supervisor, Dr. Shailly Tomar. Since I began my graduate studies, I am enlightened by her passion for science. Her painstaking guidance, unflagging interest, constructive criticism and perpetual encouragement at various stages shaped my work to be completed perfectly. Without her meticulous attention, the thesis work would not have been a success. Dr. Shailly Tomar provided me all the necessary facilities to ensure the successful voyage of my Ph. D. She has been my primary resource for getting my science questions answered and has been instrumental to crank out my thesis work. She has supported me academically and emotionally throughout my thesis. I could not have imagined having a better advisor and mentor for my Ph. D. study.

The valuable scientific studies are always the result of collaborative efforts. I endow my heartfelt thanks to Dr. Pravindra Kumar, for helping me out in the structural studies and giving me an authoritative support for the progress of my work. He is a polymath and the smartest person I know so far who encouraged me to do better. I couldn't have succeeded without his erudite support in structure biology and bioinformatics. I would like to thank him for his insightful discussions, suggestions and for all his scientific advices which provided me positive outlook during different phases of journey of my Ph. D.

I acknowledge with thanks the kind of patronage, loving inspiration and timely guidance which I received from my committee members Prof. R. P. Singh, Prof. Partha Roy and Dr. P. Gopinath. I am very fortunate to get the opportunity of having such great committee members.

Some faculty members have been very kind to extend their help at various stages of my research. I am grateful to Prof. G. S. Randhawa, Prof. Vikas Pruthi, Dr. A. K. Sharma, Dr. Naveen K. Navani, Dr. Ranjana Pathania, Dr. Harsh Chauhan, Dr. Kiran, Dr. Mandal, Dr. Deepak Sharma, Dr. Krishnanmohan Poluri, Dr. Sircar, Dr. Soma Rohtangi for their help and encouragement.

I owe my sincere thanks to Dr. Sonali Dhindwal, Dr. Megha Aggarwal, Dr. Satya Tapas, Dr. Preeti, Dr. Bibekananda Kar, Dr. Dipak Patil, Dr. Preeti Verma, Dr. Rajnikanth Sharma Dr. Manju Narwal, Dr. Monu Batra to be instrumental for helping me to get accustomed to the laboratory. I had a great time in laboratory with Dr. Sonali Dhindwal, Dr. Megha Aggarwal and Dr. Bibekananda Kar who helped me a lot as seniors and friends as well.

It is a humbling experience to acknowledge those who have, mostly out of kindness helped along the journey of my Ph. D. I thank, Ms. Pooja Kesari, Mr. Rajat Mudgal, Mr.

Madhusudhan Rao Katiki, Mr. Harvijay, Ms. Ramanjit, Ms. Benazir Fatma for the fruitful discussions. I appreciate the discussions with Dr. Sonali Dhindwal, Ms. Pooja Kesari and Mr. Rajat Mudgal who helped me whenever I was not getting positive results during my research work. I am thankful to all my labmates to facilitate the fun environment in the laboratory and also for helping me in all the possible way during my tenure.

I would like to thank Mr. Harvijay, Ms. Ramanjit, Mr. Pranav Kumar, Ms. Anjali Malik, Ms. Benazir Fatma, Mr. Akshay Pareek, Ms. Vedita Anand Singh, Ms. Gunjan Saini, Ms. Anchal Sharma, Mr. Vijay Sharma, Ms. Neha Singh, Ms. Neetu, Dr. Ravi Saini, Ms. Supreeti Mahajan and Ms. Happy Goyal for their sincere cooperation and efforts.

I must also thank Dr. Sakshi Sharma, Mr. Rajat Mudgal and Ms. Pooja Kesari for always being there with a word of encouragement or listening ear during my tough phases of Ph. D. This thesis would not exist without their love and support. I am also grateful to the Institute Sports Club as well as the Institute Hospital. Both facilitated me in maintaining physical fitness so that I could achieve my studies contained in this thesis.

Finally, and most importantly, my heartiest thank goes to my parents and all of my family members for their blessings and unconditional love. Their support and encouragement was worth more than I can express on a paper. They encouraged and supported me along the bumpy roads of my research work and inspire me to achieve the targets. They always stand firmly behind me and provide resources to carry out my studies. I would like to express my heartfelt gratitude to my family. None of this would have been possible without the love and the patience of my family. They have cherished with me every great moment and supported me whenever I needed it.

Beyond all these, I thank to almighty God for his blessings on me and keeping me strong and calm during all good and hard times throughout my life.

Finally I would like to thank to Department of Biotechnology (DBT), New Delhi India for financial assistance as JRF and SRF.

(Rajesh)

LIST OF PUBLICATIONS

1. Monu Batra, **Rajesh Sharma**, Vemika Chandra, Megha Aggarwal, Uday Agarwal, Pawan Gupta, Rajesh Pratap Singh, and Shailly Tomar. "In silico and proteomic analysis of protein methyltransferase CheR from *Bacillus subtilis*." **International journal of biological macromolecules** 77 (2015): 168-180.
2. Megha Aggarwal, **Rajesh Sharma**, Pravindra Kumar, Manmohan Parida, Shailly Tomar. "Kinetic characterization of trans-proteolytic activity of Chikungunya virus capsid protease and development of a FRET-based HTS assay". **Scientific reports** 5 (2015).
3. Monu Batra¹, **Rajesh Sharma**¹, Anjali Malik, Sonali Dhindwal, Pravindra Kumar, and Shailly Tomar. "Crystal structure of pentapeptide-independent chemotaxis receptor methyltransferase (CheR) reveals idiosyncratic structural determinants for receptor recognition." **Journal of Structural Biology** 196, no. 3 (2016): 364-374.
4. **Rajesh Sharma**, Benazir Fatma, Amrita Saha, Sailesh Bajpai, Srinivas Sistla, Paban Kumar Dash, Manmohan Parida, Pravindra Kumar, and Shailly Tomar. "Inhibition of chikungunya virus by picolinate that targets viral capsid". **Virology** 498 (2016): 265-276.
5. **Rajesh Sharma**, Pooja Kesari, Pravindra Kumar and Shailly Tomar. "Structure-function insights into chikungunya virus capsid protease: Small molecules targeting capsid hydrophobic pocket". **Virology** 515 (2018): 223-234.
6. Benazir Fatma, **Rajesh Sharma**, Pooja Kesari, Shailly Tomar. "Structure based screening and evaluation of chikungunya virus capsid protease inhibitors for drug repurposing". (Under preparation).

CONTENTS

Page No.

CANDIDATE'S DECLARATION

ABSTRACT.....	i-iv
ACKNOWLEDGEMENTS.....	v-vi
LIST OF PUBLICATIONS.....	vii
CONTENTS.....	viii-xi
LIST OF FIGURES.....	xii-xiii
LIST OF TABLES.....	xiv
LIST OF ABBREVIATIONS USED.....	xv-xvi

CHAPTER 1

REVIEW OF LITERATURE

1.1 Introduction.....	1
1.2 Alphavirus life cycle.....	3
1.3 Structure of alphavirus virion.....	5
1.4 Genome architecture.....	6
1.5 Non-structural proteins.....	7
1.6 Structural proteins.....	10
1.7 Nucleocapsid assembly.....	14
1.8 Alphavirus budding.....	16
1.9 Role of glycoproteins in virus budding.....	17
1.10 Capsid protein structure.....	17
1.11 Alphavirus CP <i>trans</i> -activity.....	20

CHAPTER 2

INHIBITION OF CHIKUNGUNYA VIRUS BY PICOLINATE THAT TARGETS VIRAL CAPSID PROTEIN

2.1 Abstract.....	22
2.2 Introduction.....	23
2.3 Material and Methods.....	26
2.3.1 3D model generation.....	26

2.3.2	Molecular Docking.....	26
2.3.3	Virus.....	27
2.3.4	Expression and Purification of CVCP protein.....	27
2.3.4.1	Expression of CVCP.....	28
2.3.4.2	Purification of CVCP.....	28
2.3.5	Circular dichroism spectroscopy.....	29
2.3.6	Isothermal titration calorimetry.....	29
2.3.7	Surface plasmon resonance.....	30
2.3.7.1	Buffer Preparation.....	30
2.3.7.2	Immobilization of CVCP protein.....	30
2.3.8	Fluorescence spectroscopy.....	31
2.3.9	Cells and virus propagation.....	31
2.3.9.1	Cytotoxicity testing.....	32
2.3.9.2	Antiviral assay.....	32
2.3.9.2.1	qRT-PCR.....	32
2.3.9.2.2	Plaque reduction assay.....	33
2.4	Results.....	33
2.4.1	3D Homology model.....	33
2.4.2	Molecular docking.....	36
2.4.3	Purification of CVCP and secondary structure analysis.....	38
2.4.4	Isothermal studies of picolinic acid binding.....	38
2.4.5	Surface plasmon resonance.....	39
2.4.5.1	CVCP protein immobilization.....	40
2.4.5.2	Kinetic analysis.....	40
2.4.6	Effect of PCA on CVCP intrinsic fluorescence.....	41
2.4.7	Cytotoxicity testing.....	42
2.4.8	Antiviral assay.....	43
2.5	Discussion.....	45
2.6	Conclusions.....	48

CHAPTER 3

STRUCTURE-FUNCTION INSIGHTS INTO CHIKUNGUNYA VIRUS CAPSID PROTEIN: SMALL MOLECULES TARGETING CAPSID HYDROPHOBIC POCKET

3.1	Abstract.....	49
3.2	Introduction.....	50
3.3	Material and Methods.....	53
3.3.1	Expression and Purification.....	53
3.3.2	Crystallization and data collection.....	53
3.3.3	Structure solution and refinement.....	54
3.3.4	Molecular modeling.....	55
3.3.5	Identification and docking of small heterocyclic molecules.....	55
3.3.6	Surface plasmon resonance (SPR)	56
3.3.7	Fluorescence spectroscopy.....	56
3.3.8	<i>In-silico</i> epitope identification.....	57
3.4	Results and Discussion.....	57
3.4.1	Structure determination and overall architecture.....	57
3.4.2	Monomer architecture.....	59
3.4.3	Substrate binding site architecture.....	60
3.4.4	Dimer architecture.....	61
3.4.5	CVCP-cdE2 glycoprotein interactions.....	64
3.4.6	Structural comparison of CVCP with other alphavirus CP's.....	68
3.4.7	Identification and docking of small heterocyclic molecules.....	70
3.4.8	Surface plasmon resonance.....	73
3.4.9	Effect of MDA and EAB on CVCP intrinsic fluorescence.....	75
3.4.10	Epitope identification.....	76
3.5	Conclusion.....	78

CHAPTER 4

CHARACTERIZATION OF TRANS-PROTEOLYTIC ACTIVITY OF CHIKUNGUNYA VIRUS CAPSID PROTEASE

4.1	Abstract.....	80
4.2	Introduction.....	81

4.3	Material and Methods.....	86
4.3.1	Expression of active CVCP.....	86
4.3.2	Purification of active CVCP.....	87
4.3.3	CVCP <i>trans</i> -protease activity.....	87
4.3.4	Effect of inhibitors on CVCP protease activity.....	89
4.4	Results and Discussion.....	89
4.4.1	CVCP expression and purification (active).....	89
4.4.2	CVCP <i>trans</i> -protease activity.....	91
4.4.3	Effect of inhibitors on CVCP protease activity.....	93
4.5	Conclusions.....	95
REFERENCES.....		97-120



LIST OF FIGURES


Figure 1.2.1:	Graphical illustration of the alphavirus transmission cycle.....	3
Figure 1.2.2:	Schematic representation of alphavirus life cycle.....	4
Figure 1.3.1:	Virion structure organization.....	5
Figure 1.4.1:	Genome organization.....	6
Figure 1.5.1:	Non-structural polyprotein processing.....	8
Figure 1.6.1:	Processing of structural polyprotein.....	11
Figure 1.6.2:	Different regions of alphavirus capsid protein.....	12
Figure 1.7.1:	Nucleocapsid assembly.....	14
Figure 1.8.1:	Schematic representation of two types of budding models.....	16
Figure 1.10.1:	Cartoon view representation of alphavirus crystal structures.....	19
Figure 2.4.1.1:	3D atomic model structure of CVCP.....	34
Figure 2.4.1.2:	Structure comparison with alphavirus capsid protein.....	34
Figure 2.4.1.3:	Multiple sequence alignment of CVCP protein with sequences of other alphavirus capsid protein (CP).....	36
Figure 2.4.2.1:	Binding analysis of docked compound into CVCP hydrophobic pocket.....	37
Figure 2.4.3.1:	SDS-PAGE analysis of chikungunya virus capsid protein (CVCP)...	38
Figure 2.4.4.1:	ITC raw and integrated data.	39
Figure 2.4.5.1:	SPR binding and kinetic analysis.....	41
Figure 2.4.6.1:	Fluorescence spectroscopy.....	42
Figure 2.4.7.1:	Apoptosis in PCA treated Vero cells.....	43
Figure 2.4.8.1:	qRT-PCR.....	44
Figure 2.4.8.2:	Plaque reduction assay.....	45
Figure 3.3.1.1:	Gel filtration chromatography and SDS-PAGE analysis of CVCP.....	53
Figure 3.3.2.1:	Crystal and Diffraction image of CVCP.....	54
Figure 3.4.2.1:	Crystal structure of CVCP.....	60
Figure 3.4.3.1:	Catalytic architecture of CVCP.....	61
Figure 3.4.4.1:	Dimer interface interactions.....	62
Figure 3.4.4.2:	Structural insights of dimer interface interactions.....	63
Figure 3.4.4.3:	Multiple sequence alignment of CVCP with CP's from other alphaviruses.....	64

Figure 3.4.5.1:	CVCP crystal structure and E1, E2 glycoprotein homology models fitted in to cryo-EM density map of VEEV.....	65
Figure 3.4.5.2:	Depiction of Helix-loop-Helix of cdE2 glycoprotein fitted in to CVCP hydrophobic pocket.....	66
Figure 3.4.5.3:	CVCP- cdE2 interactions.....	67
Figure 3.4.5.4:	Multiple sequence alignment of cytoplasmic domain of E2 glycoprotein (cdE2) from different alphaviruses.....	67
Figure 3.4.5.5:	Hydrophobic interactions of Pro404 of cdE2 with CVCP hydrophobic pocket.....	68
Figure 3.4.6.1:	Superimposition of CP's from different alphaviruses.....	70
Figure 3.4.7.1:	Binding analysis of docked (S)-(+)-Mandelic acid (MDA) into CVCP hydrophobic pocket.....	72
Figure 3.4.7.2:	Binding analysis of docked Ethyl 3-aminobenzoate (EAB) into CVCP hydrophobic pocket.....	72
Figure 3.4.8.1:	SPR binding and kinetic analysis of (S)-(+)-Mandelic acid.....	73
Figure 3.4.8.2:	SPR binding and kinetic analysis of Ethyl 3-aminobenzoate.....	74
Figure 3.4.9.1:	Intrinsic fluorescence spectroscopy analysis of MDA.....	75
Figure 3.4.9.2:	Intrinsic fluorescence spectroscopy analysis of MDA.....	75
Figure 3.4.10.1:	<i>In-silico</i> epitope identification.....	77
Figure 4.2.1:	Catalytic mechanism of serine proteases.....	84
Figure 4.2.2:	The LigPlot analysis of benzamidine binding to protease S1 pocket...	85
Figure 4.2.3:	Mechanism of inactivation of serine and cysteine proteases by halomethyl ketone.....	86
Figure 4.3.3.1:	Fluorescence resonance energy transfer (FRET) protease assay.....	88
Figure 4.4.1.1:	Schematic representation of alphavirus active and inactive protease...	90
Figure 4.4.1.2:	Expression of active CVCP.....	91
Figure 4.4.1.3:	Purification of active CVCP.....	91
Figure 4.4.2.1:	Kinetic analysis of active CVCP.....	92
Figure 4.4.3.1:	Determination of inhibition constant for BHC	94
Figure 4.4.3.2:	Determination of inhibition constant for TPCK.....	95

LIST OF TABLES

Table 1.1: Functions of non-structural proteins (nsPs).....	10
Table 1.2: Functions of structural proteins.....	13
Table 1.3: List of CP crystal structures from different members of alphavirus family	18
Table 2.1: List of the compounds with their empirical formula and molecular structure used for docking into the hydrophobic pocket of CVCP.....	27
Table 2.2: The various set parameters used for performing the Isothermal titration calorimetry experiments of dioxane and picolinic acid with purified CVCP.....	30
Table 2.3: The values of binding energy obtained from molecular docking of dioxane, proline and picolinic acid into the CVCP hydrophobic pocket....	37
Table 2.4: The thermodynamic analysis of dioxane and picolinic acid binding to CVCP as obtained from ITC.....	39
Table 3.1: Data collection and refinement statistics for CVCP.....	58
Table 3.2: CVCP intra-subdomain salt bridges.....	59
Table 3.3: Interface polar interactions between the two crystallographic monomers of CVCP.....	63
Table 3.4: The molecular docking results of dioxane, picolinic acid, (S)-(+)-Mandelic acid and Ethyl 3-aminobenzoate binding to CVCP hydrophobic pocket.....	71

LIST OF ABBREBIATIONS USED



Å	Angstrom
α	Alpha
°C	Degree Centigrade
β	Beta
μg	Microgram
μl	Microlitre
μM	Micromolar
μmole	Micromole
bp	Base pair
B-factor	Debye-Waller factor/Temperature factor
cDNA	complementary DNA
cm	Centimeter
CHIKV	Chikungunya virus
CP	Capsid Protein
CVCP	Chikungunya virus capsid protein
Da	Daltons
DNA	Deoxyribose nucleic acid
E. coli	Escherichia coli
eg.	For example
et al.	et alia
g	Gram
h	Hours
H-bond	Hydrogen bond
IITR	Indian Institute of Technology Roorkee
IPTG	Isopropyl β -D-thiogalactoside
ITC	Isothermal titration calorimetry
K	Kelvin
kDa	Kilo Daltons
m	Meter
M	Molar
mg	milligram
min	Minute
ml	millilitre

mm	millimeter
mM	millimolar
NaCl	Sodium Chloride
NCBI	National Center for Biotechnology Information
NI-NTA	Nickel-nitrilotriacetic acid
nm	Nanometer
PCR	Polymerase chain reaction
PDB	Protein Data Bank
R-factor	Residual-factor
RNA	Ribose nucleic acid
rpm	Revolutions per minute
s	Seconds
SPR	Surface plasmon resonance
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
v/v	volume /volume



Chapter 1

Review of Literature

1.1 Introduction

Alphaviruses are a group of spherical, enveloped, single stranded positive sense RNA viruses. Alphavirus genus belongs to *Togaviridae* family and includes many medically relevant human and animal viruses like Ross River virus (RRV), Venezuelan Equine Encephalitis virus (VEEV), Chikungunya virus (CHIKV), Western Equine Encephalitis virus (WEEV) etc. The genus Alphavirus contains at least 29 members that are capable to infect several vertebrates like humans, horses, rodents, birds, fish and birds. These members have been currently classified into eight complexes based on antigenic and genetic similarities (Weaver et al, 2005; Powers et al., 2001). Alphaviruses are classified into New and Old World viruses based on the mechanism employed for shutting the host transcription off, disease presentation and mortality rate. Old World viruses like Aura virus, Sindbis virus (SINV), Chikungunya virus (CHIKV) and Semliki Forest virus (SFV) etc. utilize nsp2 protein to down regulate host cell transcription, cause arthralgia and have low mortality rate. New World viruses like VEEV, WEEV etc. utilize capsid protein to down regulate host cell transcription, cause encephalitis and have high mortality rate (Garmashova et al., 2007; Hahn et al., 1988).

The cryo-electron microscopy (cryo-EM) and crystallographic studies from various alphaviruses provide crucial details about the distribution, structures and organization of virion. The size of alphavirus genome is approximately ~11.7 kb, capped at 5' end and poly-adenylated at 3' end. The nucleocapsid core (NC) is formed of RNA genome encapsidated by around 240 copies of capsid protein (Baron et al., 1996; Vasiljeva et al., 2000). The mature alphavirus particle has envelop derived from host cell membrane, embedded with 80 spikes in T=4 icosahedral symmetry. Each spike is made up of trimers of E1 and E2 glycoprotein heterodimers (Mukhopadhyay et al., 2006; Soonsawad et al., 2010; Strauss and Strauss, 1994). The N-terminal two-third and C-terminal one-third section of alphavirus genome translate into non-structural and structural proteins, respectively. The non-structural polyprotein is produced by the translation of genomic 49S RNA that via minus strand RNA intermediate also transcribes the 26S sub-genomic RNA. Upon translation, the 26S sub-genomic RNA produced structural polyprotein. The processing of non-structural polyprotein produced nsp1, nsp2, nsp3 and nsp4 proteins. The structural polyprotein comprises of CP at the N-terminus tailed by E3, E2, 6K and E1 proteins.

The alphavirus capsid protein (CP) has multiple functions. It is involved in autoproteolysis, nucleocapsid core formation, capsid-glycoprotein interactions, host transcription down regulation and contain signals for nuclear and cytoplasmic trafficking. It

comprises of two domains, RNA binding N-terminal domain and the C-terminal protease domain (Choi et al., 1991; Melancon and Garoff, 1987; Strauss and Strauss, 1994). The N-terminal domain is responsible for RNA genome encapsidation, host transcription down regulation and capsid-capsid dimerization. The C-terminal protease domain possesses *cis*-autoproteolytic activity that cleaves itself at W/S scissile bond once in virus life cycle. After the cleavage the terminal tryptophan residue of CP binds to the catalytic site and inhibits further CP protease activity. The C-terminus protease domain of CP contains a conserved hydrophobic pocket at the surface, a site for capsid-glycoprotein interactions. The CP-glycoprotein interactions are pre-requisite for alphavirus budding process. The key role of alphavirus CP in virus life cycle makes it a potential drug target.

In the present study, the CP from CHIKV has been studied for its protease activity and the protein-protein interactions (PPI's) of CP with glycoproteins. We explored new approaches for development of antivirals against alphaviruses. The crystal structure of chikungunya virus capsid protease (CVCP) has been determined and detailed structural analysis has been carried out. In the absence of crystal structures of CHIKV E1 and E2, models were made utilizing the cryo-EM structure of VEEV as template. The CVCP crystal structure and the generated E1 and E2 models were fitted in to the EM density map of VEEV to study the PPI's between CP and glycoproteins. Structural analysis has revealed the hydrophobic pocket of CVCP is involved in the CP-glycoproteins interaction. Previous studies have suggested that dioxane binding in the CP hydrophobic pocket mimics the pyrrolidine ring of the conserved proline of cytoplasmic domain of E2 glycoprotein (cdE2) (Aggarwal et al., 2012). This suggested that heterocyclic ring compounds similar to dioxane and their derivatives could bind in the hydrophobic pocket of CP and block CP-cdE2 interactions. Based on this critical interaction, the small heterocyclic molecules [(Picolinic acid (PCA), (S)-(+)-Mandelic acid (MDA) and Ethyl 3-aminobenzoate (EAB)] were identified. The binding of small heterocyclic molecules to CVCP hydrophobic pocket was analyzed by molecular docking, isothermal titration calorimetry (ITC), fluorescence and surface plasmon resonance (SPR). Additionally, the antiviral properties of PCA against CHIKV have been studied.

In 2008, Morillas et al., reported the esterase activity of SFV truncated CP by deleting the C-terminal seven residues. The CVCP construct by deleting the two residues from the C-terminal end of CP has been cloned, expressed and purified. The truncated CVCP (active) has been characterized for *in vitro trans*-protease activity. A fluorescence resonance energy transfer (FRET) based CVCP *trans*-protease assay has been developed. This protease assay can be used for high throughput screening (HTS) of inhibitors against the alphavirus capsid protease.

1.2 Alphavirus life cycle

The Alphavirus genus includes many medically relevant human and animal viruses like RRV, CHIKV, VEEV, WEEV etc. The viral replication takes place inside the mosquito and spread further through the salivary secretion of the vector. These are transmitted to humans by the bite of virus-carrying Aedes mosquito and the transmission follows the mosquito-vertebrate-mosquito cycles. Humans can catch the infection either through mosquito bite or through other vertebrate hosts. The graphical representation of the alphavirus transmission is shown in Fig. 1.2.1.

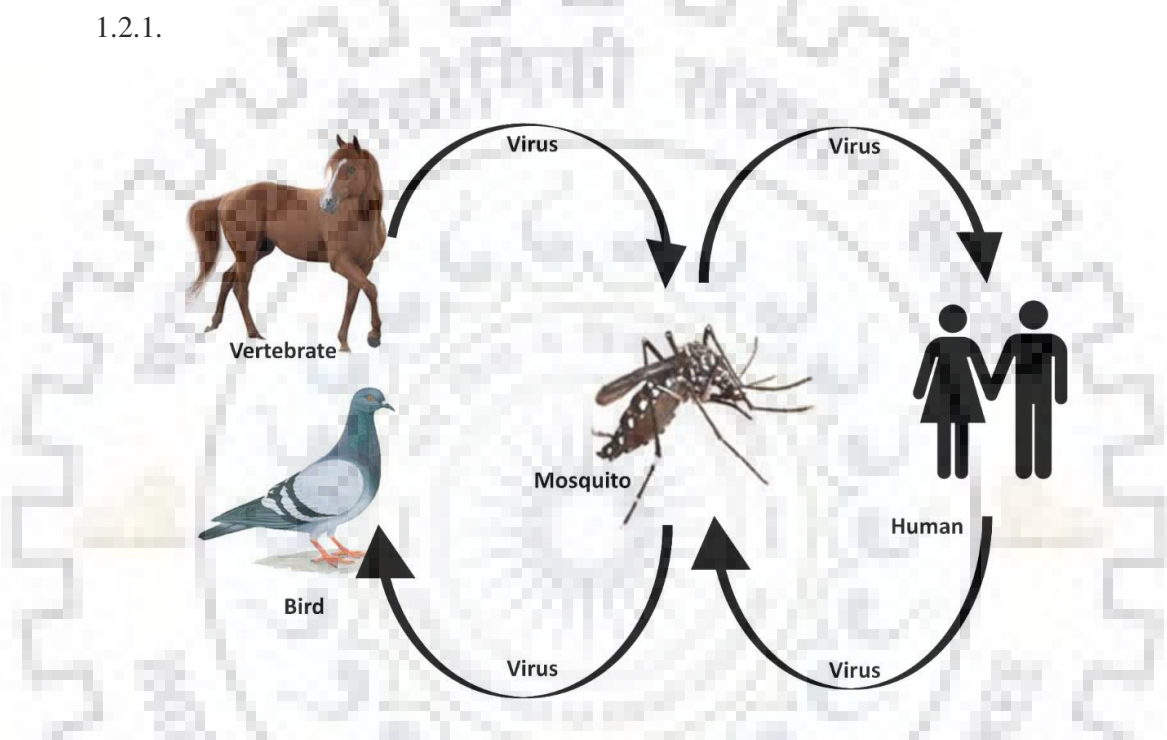


Figure 1.2.1: Graphical illustration of the alphavirus transmission cycle (Schmaljohn et al., 1996).

The life cycle initiates with the engagement of virus to the host cell membrane via specific receptors. Alphaviruses employ a variety of host cell receptors (laminin, heparin sulphate, integrin, C-type lectins DC-SIGN and L-SIGN etc.) for entry in to host cell (Klimstra et al., 2003; 1998). The schematic representation of alphavirus life cycle is given in Fig. 1.2.2. The alphavirus E2 glycoprotein is mainly responsible for interactions with receptor followed by entry of virus through receptor mediated endocytosis. The acidification of endosome due to presence of vacuolar ATPase facilitate the fusion of virion particle to the endosomal membrane (Justman et al., 1993; Wahlberg & Garoff, 1992; Mellman et al., 1986; Kielian et.al., 1985; 1995; 2002). This drop in pH induces conformational changes in E1-E2 glycoproteins which results in dissociation of E1-E2 heterodimer and subsequent E1 fusion peptide insertion to endosomal membrane (Kielian et.al., 1085; Kondor-Koch et al., 1983). Before the drop in endosomal pH,

the fusion peptide was concealed behind the E1-E2 heterodimer and was not accessible for interaction with endosomal membrane. These sequence of events results in release of nucleocapsid core (NC) in to the host cytoplasm (Singh & Helenius, 1992). The disassembly of virus NC in the cytoplasm release the encapsidated genomic RNA for translation.

The genomic RNA acts as mRNA and translates from two ORF's to produce non-structural (P1234) and structural polyproteins. The *cis*-cleavage of P1234 between nsp3 and nsp4 generate P123 and nsp4. Further processing of P123 yields nsp1, nsp2 and nsp3 proteins. These non-structural proteins along with some unknown factors form replication complex. The translation of subgenomic 26S RNA produces structural polyprotein which upon processing generate CP, E3, E2, 6K and E1 structural proteins.

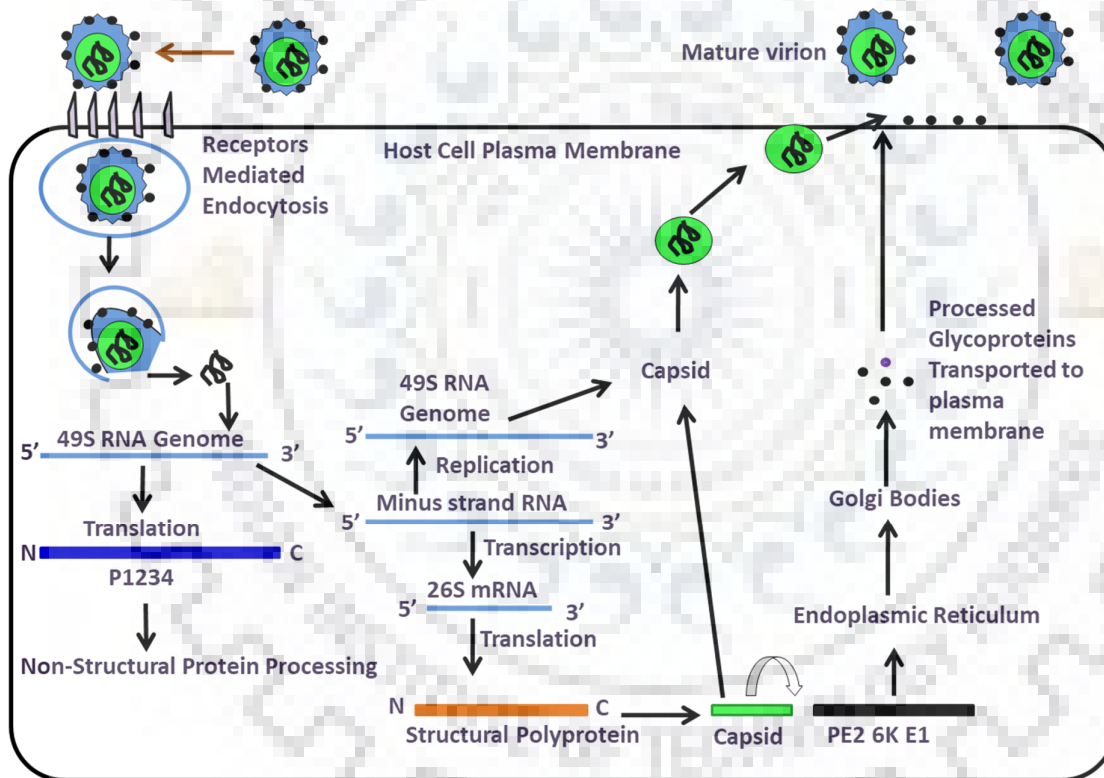


Figure 1.2.2: Schematic representation of alphavirus life cycle. The virion enters the cell via receptor mediated endocytosis. A drop in endosomal pH induces conformational changes in the E1-E2 glycoproteins which results in disassembly of nucleocapsid core. The released genomic RNA translate into non-structural polyprotein which upon processing produce four non-structural proteins. The 26S subgenomic RNA translates to form structural polyprotein comprising of structural proteins in the order of CP, E3, E2, 6K and E1. The CP cleaves itself from rest of the structural polyprotein and facilitate the further processing of N-terminal structural polyprotein in endoplasmic reticulum and Golgi apparatus. The CP interacts with the genomic RNA and form nucleocapsid core. The interaction of CP with the glycoproteins leads to budding of the virus particle from the surface of host cell (Jose et al., 2009).

The CP has a chymotrypsin-like fold and possesses *cis*-autoproteolytic activity (Aliperti et al., 1978; Hahn et al., 1990; Aggarwal et al., 2014). The CP cleaves itself at W/S scissile bond and separates out from rest of the structural polyprotein. After the cleavage of CP, the N-terminus of the remaining structural polyprotein now contains signal sequence responsible for translocation of structural polyprotein to endoplasmic reticulum and Golgi for further processing and post-translational modifications. The CP encapsidates the genomic RNA and results in the nucleocapsid assembly and thus formation of nucleocapsid core (Warrier et al., 2008). The CP is also involved in interactions with the surface glycoproteins, required for budding of virus from host cell surface. The interactions of cdE2 with the hydrophobic pocket of CP is pre-requisite for budding of mature virions from the surface of host cell. During the translocation of the pE2–E1 complex from Golgi to cellular membrane, pE2 is sliced by furin to produce E2 and E3. For fresh infection to host cell, the cleavage of pE2 is essential as it involves in virus entry followed by fusion peptide activation (Ryman et al., 2004; 2007).

1.3 Structure of alphavirus virion

The mature alphavirus virion is about 700 Å in diameter with icosahedral symmetry and host derived envelop. The cryo-electron microscopy studies of whole virion and crystallography studies of component proteins have been extensively used to carry out the structural studies of alphaviruses (Pletnev et al., 2001; Mancini et al., 2000; Cheng et al., 1995). The cryo-electron microscopic (cryo-EM) structures for SFV, VEEV, AURAV, RRV, SINV and CHIKV are available at high resolutions (Cheng et al., 1995; Zhang et al., 2002; 2011; Mukhopadhyay et al., 2006; Mancini et al., 2000; Sun et al., 2013). All the members of alphavirus genus share similar structural features. The molecular mass and density of alphavirus particle are 5.2×10^6 Da and 1.22 g/cm^3 , respectively (Paredes et al., 1992., Cheng et al., 1995). The structural organization of Sindbis virus is represented in Fig. 1.3.1.

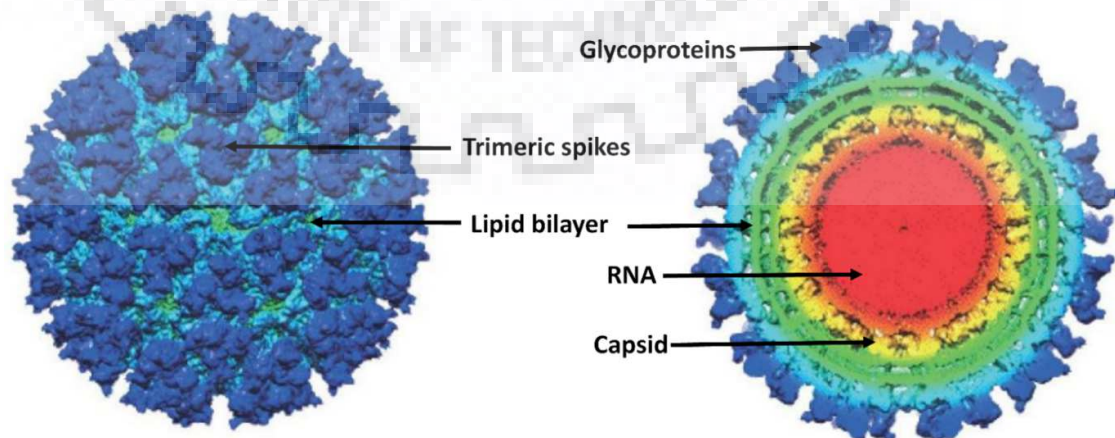


Figure 1.3.1: Virion structure organization. In the left panel, icosahedral two fold axis of surface-shaded view of Sindbis virus at 9 Å is shown. The trimeric spikes and lipid bilayer are represented by blue and green color, respectively. Right panel depicts the cross-section of the virus particle representing the surface glycoproteins, lipid bilayer, capsid protein and RNA in blue, green, yellow and red color, respectively. (Jose et al., 2010).

The mature virion surrounded by two concentric shells consists of 240 copies of each capsid, E1 and E2 protein. The outer shell comprises of E1 and E2 glycoproteins and inner shell comprises of CP (Cheng et al., 1995; Strauss and Strauss, 1994). The space between the two concentric shells is filled with the host derived lipid bilayer through which the transmembrane helix of the glycoproteins penetrates (Mancini et al., 2000; Tang et al., 2011). The host derived envelope is embedded with 80 spikes in T=4 icosahedral symmetry (Paredes et al., 1993; Fuller, 1987; Vogel et al., 1986; von Bonsdorff & Harrison, 1975). Each spike is made up of trimers of E1 and E2 (surface glycoproteins) heterodimers (Mukhopadhyay et al., 2006; Soonsawad et al., 2010; Strauss and Strauss, 1994; Zhang et al., 2002). The interaction of N-terminal domain of CP with the genomic RNA leads to RNA encapsidation and formation of nucleocapsid core. Additionally, in Aura virus, the CP is also involved in encapsidation of subgenomic RNA (Rumenapf et al., 1995).

1.4 Genome architecture

Alphavirus genome consists of positive sense single stranded RNA, approximately ~11.7 kb size; capped at 5' end and poly-adenylated tail at 3' end. The genomic RNA comprises of two open reading frames (ORF's), to produce structural and non-structural polyprotein. The N-terminal two third of the genome is responsible for translation of non-structural proteins (nsp1-nsp4) which play key role in virus life cycle such as polyprotein processing, replication, RNA transcription and capping. The C-terminal one third part of genome encodes for structural polyprotein that contains capsid and glycoproteins. The genome organization of alphavirus is shown in Fig. 1.4.1.

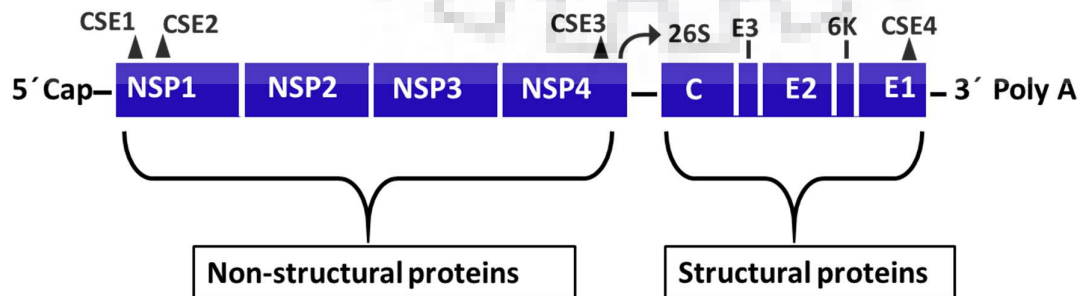


Figure 1.4.1: Genome organization. The alphavirus RNA genome has of 5' cap and 3' poly-A tail. The two third region at N-terminus of the alphavirus genome translates into non-structural polyprotein, which upon processing yield non-structural proteins (nsP1-nsP4). The one third region at C-terminus of genome translates into structural polyprotein, which undergoes processing to produce structural proteins in the order of CP, E3, E2, 6K and E1. Additionally, alphavirus genome contains four conserved sequence elements (CSEs) which help in its transcription (Strauss et al., 1986; Rupp et al., 2015).

The alphavirus genome has four conserved sequence elements (CSE1, CSE2, CSE3 and CSE4) that are known to interact with viral or host proteins. The interactions of CSE's with viral or host proteins modulate alphavirus replication process. The CSE1 present at the 5' end of the genome and contains approximately 44 nucleotides of viral genome and acts as promoter for plus-strand RNA synthesis (Ou et al., 1982). The CSE1 region is complementary to minus-strand viral RNA that acts as template for synthesis of plus-strand RNA. The CSE2 present near the 5' end is 51 nucleotides long and acts as promoter for minus strand RNA production (Frolov et al., 2001). The CSE3 is 24 nucleotides long and present at the junction flanked by coding sequence of the structural and non-structural proteins. The presence of CSE3 aids in efficient transcription of subgenomic RNA (Pushko et al., 1997). The fourth CSE (CSE4) is 19 nucleotides long and is positioned at the 3' end of the genome is supposed to act as co-promoter for synthesis of minus strand RNA (Hardy, 2006; Kuhn et al., 1990).

1.5 Non-structural proteins

The N-terminal two third region of the RNA genome encodes for non-structural polyprotein which upon processing yields nsP1, nsP2, nsP3 and nsP4 proteins. These proteins are crucial for genomic RNA replication and transcription of sub-genomic RNA. The graphical representation of processing of non-structural polyprotein is shown in Fig. 1.5.1.

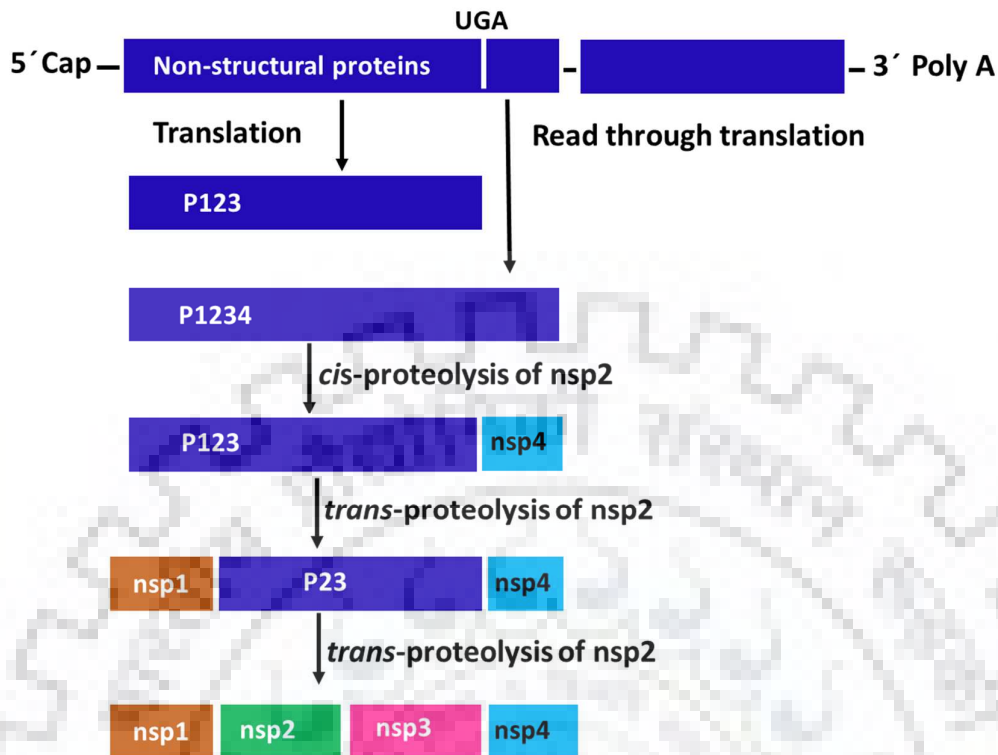


Figure 1.5.1: Non-structural polyprotein processing. The two third region of the genome at the N-terminus translates into non-structural polyprotein. In the presence of opal codon (UGA), translation of P123 takes place. While the read through of the opal codon yields P1234 which upon *cis* cleavage by the nsP2 protease produce nsp4 protein. Subsequently, the *trans*-protease activity of nsp2 is involved in processing of P123 in to nsP1, nsP2 and nsP3 (Hardy et al., 1988; 1989).

The protease activity of alphavirus nsp2 is critical for non-structural polyprotein processing. Upon infection the released RNA genome gets translated into one or two non-structural polyproteins (P123 or P1234) depending upon the presence or absence of opal codon between nsp3 and nsp4. The nsp2 protease possesses both *cis* as well as *trans*-protease activity (Hardy et al., 1989). At the infection stage, nsp2 protease *cis* activity is required for the cleavage of P1234 into P123 and nsp4. Further the nsp2 protease activity in *trans* is required for release of nsp1 from P123. Finally, the *trans*-cleavage activity of nsp2 at the cleavage site between nsp2 and nsp3 releases nsp2 and nsp3 proteins. Thus, the processing of non-structural polyprotein by proteolytic activity of nsp2 protein yields nsP1, nsP2, nsP3 and nsP4 proteins (Hardy & Strauss, 1989; 1988). These nsPs along with some unknown factors form replication complex. The following are the functions of alphavirus non-structural proteins:

NSP1: The nsp1 is a membrane associated protein that aids in anchoring of replication complex to the cellular membrane (Peranen et al., 1995). The nsP1 is also known as the capping enzyme as it exhibits guanylyl transferase and guanine-7-methyltransferase activities vital for

genomic and the sub-genomic viral RNA capping (Ahola & Kääriäinen, 1995; Mi & Stollar, 1991). Apart from this, it is also involved in maintenance of replicative intermediates of minus-strand RNA (Wang et al., 1991).

NSP2: The alphavirus nsp2 possesses multiple functions essential for processing of replication enzymes and viral replication. The N-terminal domain possesses helicase, nucleoside triphosphatase and RNA triphosphatase activities (Vasiljeva et al., 2001; Vasiljeva et al., 2000; de Cedron et al., 1999; Rikkonen et al., 1994). During replication, the unwinding of RNA genome is mediated by nsp2 helicase activity. The C-terminal domain contain a cysteine protease and a non-functional methyltransferase (MTase) (Russo et al., 2006). The nsp2 protein contain both *cis* and *trans* protease activities required for the processing of non-structural polyprotein into individual nsPs.

NSP3: The nsP3 is involved in synthesis of sub-genomic and minus-strand RNA (Lemm et al., 1994; Hahn et al., 1989; Shirako et al., 1994). The nsP3 possess RNA-binding and ADP-ribose 1-phosphate phosphatase activity. The nsp3 ADP-ribose 1-phosphate phosphatase activity is believed to be associated with induction of host cell apoptosis (Malet et al., 2009; Park & Griffin, 2009). The hyperphosphorylated form of nsp3 was suggested to involve in RNA synthesis. (Vihinen et al., 2000; Vihinen et al., 2001).

NSP4: The non-structural protein nsP4 exhibits RNA-dependent RNA-polymerase (RdRp) and terminal adenylyltransferase (TATase) activities and plays the key role in virus genome replication (Tomar et al., 2006; Hahn et al., 1989). Like other viral RdRp's, the GDD motif required for RdRp activity is also observed in alphavirus nsP4 protein (Kamer et al., 1984; Hahn et al., 1989). It is also involved in interaction with nsP1 during the formation of replication complex (Shirako et al., 2000). *In vitro* TATase activity of nsp4, which is postulated to be involved in the synthesis of poly (A) tail at the 3' end of viral genome has also been reported (Tomar et al., 2006).

Table 1.1: Functions of non-structural proteins.

Viral protein	Size (kDa)	Functions
nsP1	61	Guanylyl transferase and guanine-7-methyltransferase activities are responsible for capping of viral genomic and subgenomic RNAs. It also involved in interaction with nsP4
nsP2	90	Multifunctional protein, has two domains: N-terminal domain possesses helicase, nucleoside triphosphatase and RNA triphosphatase activities. The C-terminal domain contains non-functional methyltransferase (MTase) domain. Nsp2 also contains both <i>cis</i> and <i>trans</i> protease activities required for the processing of non-structural polyprotein into nsP1, nsP2, nsP3 and nsp4 protein.
nsP3	60	The nsP3 possesses RNA-binding and ADP-ribose 1-phosphate phosphatase activity. Apoptosis induction in host cell.
nsP4	69	RNA-dependent RNA-polymerase (RdRp) activity, virus replication, interaction with nsP1 and TATase activity.

1.6 Structural proteins

The translation of C-terminal one-third region of the RNA genome forms structural polyprotein, which upon processing by CP and furin protease, yields structural proteins in order of CP-E3-E2-6K-E1. The N-terminus of CP contains a translational enhancer, which aids the translation of subgenomic RNA to produce structural polyprotein (Frolov & Schlesinger, 1994). Structural proteins are required for budding, nucleocapsid assembly, virion entry into the host cells, shutting the host transcription off, and nucleus-cytoplasmic trafficking etc. The graphical representation of processing of structural polyprotein is shown in Fig. 1.6.1

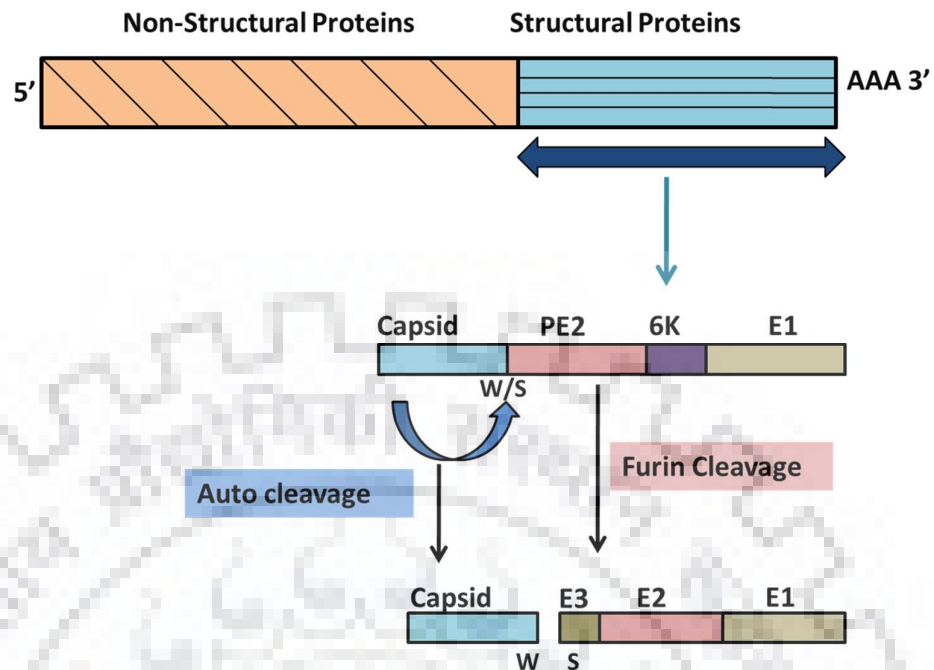


Figure 1.6.1: Processing of structural polyprotein. The carboxyl terminus one-third region of the genomic RNA encodes for the structural polyprotein. This subgenomic RNA translates to form structural polyprotein which gets processed by CP and furin to produce the CP and the glycoproteins. (Melancon et al., 1987)

The N-terminus of the structural polyprotein contains CP that has chymotrypsin like fold. It is divided into RNA binding N-terminal domain and C-terminal protease domain. The alphavirus CP (SINV) is further divided into three regions: region I (residues range: 1 to 80), region II (residues range: 81-113) and region III (residues range: 114-264) (Fig. 1.6.2). The regions I and II are part of N-terminal domain of CP and involved in binding with genomic RNA followed by its encapsidation (Hong et al., 2006). The region III is a part of C-terminal domain of CP and is responsible for protease activity of CP.

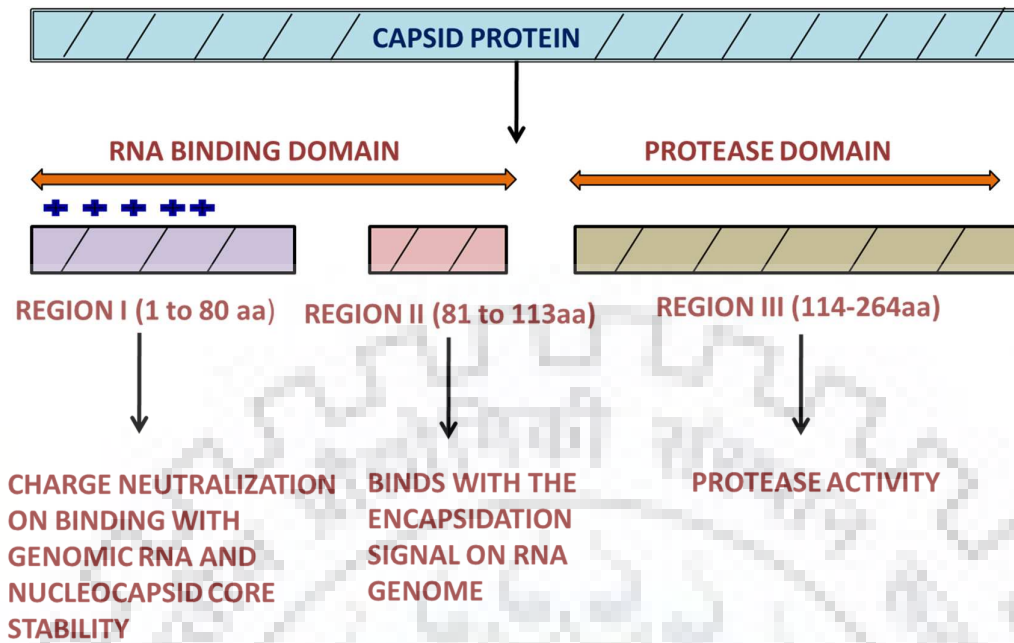


Figure 1.6.2: Different regions of Alphavirus capsid protein (SCP residues range: 1-264)

The CP has *cis*-proteolytic activity that cleaves itself from the nascent structural polyprotein precursor (Choi et al., 1991; Tong et al., 1993; Hahn & Strauss, 1990; Hahn et al., 1985; Aliperti & Schlesinger, 1978). The rest of the polyprotein now has a signal peptide sequence at the N-terminus which helps in its translocation to the endoplasmic reticulum (ER) and Golgi bodies (Garoff et al., 1978; Wirth et al., 1977). The pE2 protein undergoes post-translational modifications (palmitoylation) before its translocation from ER to *cis*-golgi. Palmitoylation of conserved cysteine residues of cE2 helps in anchoring and orienting of cE2 to the inner side of membrane that facilitate the interactions with CP (Hong et al., 2006; Jose et al., 2012). Prior to the translocation of pE2-E1 heterodimer complex from *trans*-Golgi to plasma membrane, it undergoes further processing by furin cleavage that cleaves pE2 to E2 and E3 (Jain et al., 1991; Garoff & Simons, 1974). E3 has an important role in heterodimer stabilization and its transit towards the cellular membrane for budding (Lobigs et al., 1990). To generate the infectious particles, the cleavage of pE2 is required as it involves in virus entry and peptide fusion activation (Ryman et al., 2004; 2007).

The E1-E2 heterodimer, self-assemble on the surface of the virus to form trimeric spikes (Mukhopadhyay et al., 2006; Soonsawad et al., 2010; Strauss and Strauss, 1994). The mature virion envelope is embedded with 80 trimeric spikes which are organized in T=4 icosahedral symmetry (von Bonsdorff & Harrison, 1978). The virus enters the host cell via receptor mediated endocytosis that involve the interactions of E2 glycoprotein with the surface receptors. The cryo-

electron microscopy studies have found N-linked glycosylation sites on the E2 glycoprotein responsible for binding to heparin sulphate (Knight et al., 2009; Ryman et al., 2007).

6K is a small structural protein having the size of 6000 Da. Small number of copies (up to 30) are incorporated into virions (Gaedigk-Nitschko et al., 1990; Lusa et al., 1991). It is an important structural component and undergoes palmitoylation at conserved cysteine residues, crucial for formation of infectious particle (Liljestrom et al., 1991; Gaedigk-Nitschko & Schlesinger, 1991; Gaedigk-Nitschko et al., 1990). The mutation in the cysteine residues that undergoes palmitoylation results in abnormal virion particle with reduced infectivity. In mammalian cells, 6K protein is associated with membrane modification by forming cation-selective ion channels (Madan et al., 2005; 2008; Sanz et al., 2003; Melton et al., 2002; Sanz et al., 1994). Therefore, the 6K protein can be considered as viroporin, involve in membrane permeability, glycoprotein trafficking, apoptosis and viral release (Madan et al., 2008; Gonzalez & Carrasco, 2003). It is also involved in translocation of pE2-E1 heterodimer at the site of virus assembly (Lusa et al., 1991).

The E1 is an envelope glycoprotein involved in the formation of icosahedral shell of the virus particle and membrane fusion during entry of virus in host cell. The E1 and E proteins of alphavirus and flavivirus, respectively share similar overall fold (Strauss & Strauss, 2001; Rey et al., 1995). The virion enters the cell via interaction of E2 protein with host receptors i.e. receptor mediated endocytosis. The drop in endosomal pH exposes the previously hidden E1 fusion peptide that is now available for interaction with endosomal membrane for NC disassembly the release of viral RNA. Additionally, the N-linked glycosylation sites are also reported in the E1 glycoprotein (Pletnev et al., 2001).

Table 1.2: Functions of Structural proteins.

Viral protein	Size (kDa)	Function
Capsid	30	Multifunctional protein comprises of two domains: N and C-terminal. The N-terminal domain is involved in encapsidation of genomic RNA, host transcription shut off, nucleus trafficking through NLS (nuclear localization signal). The C-terminal domain is involved in autoprotease activity, processing of structural polyprotein, cytoplasmic trafficking through NES and interaction with glycoproteins that results in virus budding.
E3	7	E3 has a critical role in heterodimer stabilization and its translocation towards the cellular membrane for budding and provide pH protection during viral assembly.

E2	47	Responsible for interactions with host cell receptors and subsequent entry through receptor mediated endocytosis. The CP-E2 glycoprotein interactions leads to budding of virions from cell surface.
6K	6	Acts as viroporin, involve in membrane permeability, translocation of pE2-E1 heterodimer at the site of viral assembly and viral release.
E1	48	Involved in fusion with endosomal membrane that leads to release of RNA genome into cytoplasm.
TF	8	TF is also called as Transframe protein. It has ion-channel activity analogous to that of 6K and plays a role in alphavirus assembly

1.7 Nucleocapsid assembly

The alphavirus genomic RNA in complex with 240 copies of capsid protein forms Nucleocapsid core (NC). Isolated NCs from the infected host cell shows a sedimentation coefficient of 140S. Under *in vitro* conditions, the purified capsids in presence of RNA, have the tendency to undergo self-assembly and form NC similar to the native one. Corresponding to the surface glycoprotein (E1-E2) arrangement, NCs also exhibit T=4 icosahedral symmetry called as capsomeres (Fig. 1.7.1).

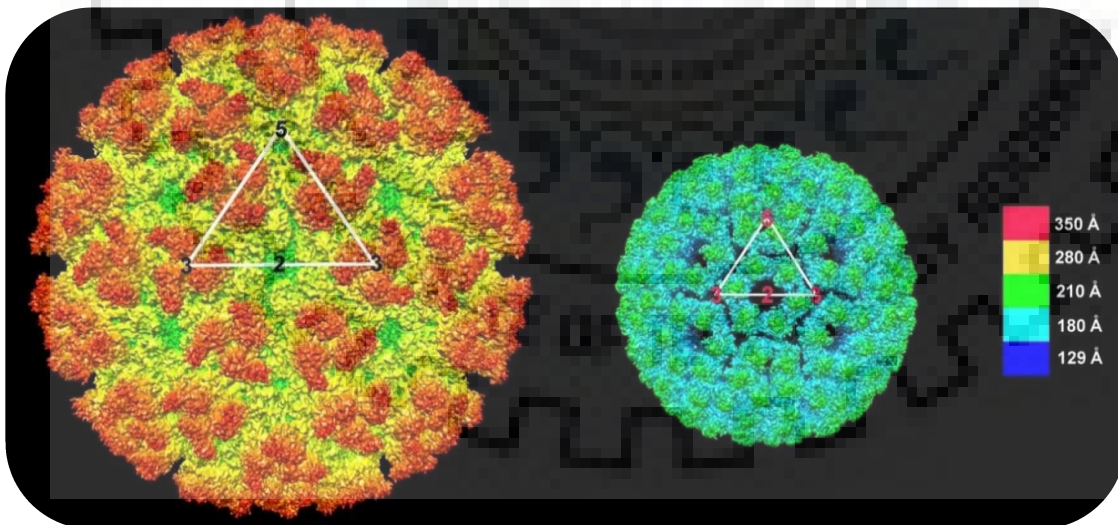


Figure 1.7.1: Nucleocapsid assembly. CHIKV cryo-EM structure (5.3 Å resolution) representing ectodomain (left) and nucleocapsid (right) as shaded surface. One icosahedral asymmetric unit is represented by white triangle (Sun et al., 2013).

Alphavirus CP has two major domains, the N-terminal domain being less conserved and highly disordered with high degree positive charge (rich in basic amino acids), is responsible for RNA genome encapsidation (Owen & Kuhn 1996; Forsell et al., 1995; 1996; Geigenmüller-Gnirke et al., 1993). Apart from RNA genome encapsidation, it participates in shutting host transcription off and dimerization of CP (Owen & Kuhn 1996). Intrinsically disordered regions plays a critical role in functional activity of the protein by exhibiting interactions with other proteins (Uversky, 2000; 2002). The interactions of disordered N-terminal domain of CP with RNA genome followed by capsid dimerization, ultimately leads to nucleocapsid core formation (Perera et al., 2001; Owen & Kuhn, 1996). Alike, the disordered N-terminal domain of alphavirus nsP4 protein also participate in PPI's (Rupp et al., 2011; Tomar et al., 2006). The importance of alphavirus CP in key biological functions makes it a potential drug target.

The alphavirus NC assembly is a multistep process. The binding of nucleic acid to CP dimer is an early step in the NC assembly process. (Tellinghuisen et al., 1999; 2001). In SINV, the region of N-terminal domain of CP from residues 1 to 80 is involved in non-specific binding to RNA genome (Perera et al., 2003). The CP region spanning from residues 81 to 113 is associated with specific binding to RNA genome encapsidation signal (Linger et al., 2004; Owen & Kuhn, 1996). Extensive *in vitro* and *in vivo* studies have been carried out to confirm the role of these regions (Owen & Kuhn, 1996; Wengler et al., 1982; 1992). Mutational studies in the region 76-107 of CP have reported the complete loss of binding of CP to RNA (Weiss et al., 1989; 1994). The stretch of 18 residues long region from 38-55 (SINV) of the N-terminal CP is projected to fold as a α -helix (helix I) (Perera et al., 2001). The helix I contains three conserved Leu residues which form the leucine zipper. The substitution of leucine residues or partial deletion of helix I has been associated with reduced viral replication. The helix I is suggested to be involved in inter-capsomeric interaction and assembly intermediates stabilization in alphaviruses. *In vitro* studies have suggested the role of residues 99, 103 and 105 in capsid dimer formation through coiled-coil interactions of helix I from each capsid monomer (Perera et al., 2003; Warriar et al., 2008). In the early steps of capsid assembly, coiled-coil interactions suggested to exhibit a regulatory role (Hong et al., 2006).

The C-terminal domain of alphavirus CP is a chymotrypsin-like serine protease with high degree of sequence and structural conservation. It possesses *cis*-proteolytic activity that cleaves at the W/S scissile bond and separates CP from the structural polyprotein (Choi et al., 1991; Melancon and Garoff, 1987). The catalytic triad is highly conserved among all the serine proteases and consists of Ser, Asp and His residues (Choi et al., 1996; Tong et al., 1993; Choi et

al., 1991; Hahn & Strauss, 1990; Melancon & Garoff, 1987). Additionally, the protease domain of alphavirus CP contains a conserved hydrophobic pocket on the surface, a site for protein–protein interaction with the glycoproteins. The interactions of CP hydrophobic pocket with the cytoplasmic domain of E2 glycoprotein is pre-requisite for the alphavirus budding (Wilkinson et al., 2005; Owen & Kuhn, 1997; Skoging et al., 1998; 2000; Suomalainen et al., 1992). Thus, the CP is a potential target for developing antivirals against alphaviral diseases.

1.7 Alphavirus budding

Budding of mature virions from the cellular membrane is the last stage of alphavirus lifecycle. Two school of thoughts have been proposed for the process of alphaviral budding (Fig. 1.8.1). First one suggests that prior to budding process through binding of CP to the glycoproteins, the formation of nucleocapsid assembly is pre-requisite. While the second model suggests that glycoproteins interact with the poorly organized capsid-RNA complex for budding process to takes place. The first model is described originally for the alphavirus budding process (Brown, 1980; Garoff & Simons, 1974). As per this model, for budding to take place, the nucleocapsid assembly having T=4 icosahedral symmetry interacts with surface glycoproteins (Acheson & Tamm, 1967).

Mutational studies in SFV, in which RNA binding and the assembly forming region were missing undergoes budding even without NC assembly (Forsell et al., 2000). In this study, a deletion mutant devoid of RNA-binding, the interdomain peptide and leucine zipper regions was prepared. Upon infection of SFV to BHK cells, the virus particles similar to wild type have been observed. Structural analysis of cryo-EM structure of this deletion mutant shows the intact glycoprotein well organized in to T=4 icosahedral symmetry. Thus, the exact mechanism and the driving force for budding are not clear yet.

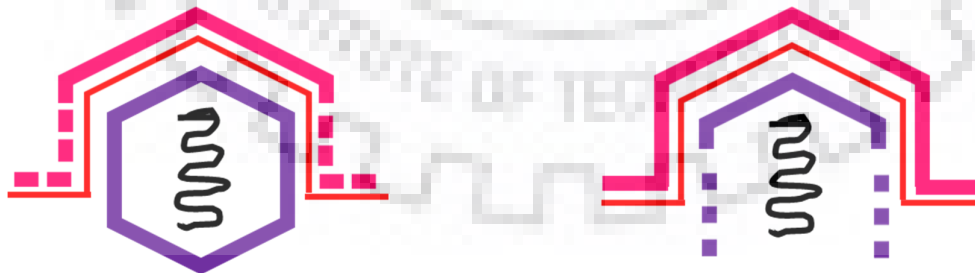


Figure 1.8.1: Schematic representation of two types of budding models. The left panel shows the nucleocapsid (violet) is organized first and then binds to the glycoproteins (magenta) that leads to budding. After the organization of nucleocapsid in icosahedral symmetry, it enforces the organization of glycoproteins. The right panel depicts the glycoprotein directed budding. In this model, the glycoproteins are organized into T=4 icosahedral symmetry without the assembled nucleocapsid. The budding takes place in the absence of properly organized capsid-RNA complex (Garoff et al., 2004).

1.9 Role of glycoproteins in virus budding

The budding of virions from the surface of plasma membrane involves the interaction of capsid and glycoprotein through the cdE2 (Strauss et al, 1995; Zhao & Garoff, 1992; 1994). During the translocation of glycoproteins heterodimer from Golgi to surface of cellular membrane, it undergoes post-translational modifications. The conserved cysteine residues (396, 416 and 417) of cdE2 undergo palmitoylation that assist in anchoring and correct orientation of cdE2 to surface membrane for enabling interactions with CP (Hong et al., 2006; Jose et al., 2012). The role of Tyr400, a residue of cytoplasmic tail of E2 in SINV and SFV has been demonstrated by mutational studies. For proper virus dissemination, this residue is essential and can be substituted only by the hydrophobic residue (Zhao et al., 1994; Gaedigk-Nitschko & Schlesinger, 1991). Another study, has demonstrated the role of phosphorylation and dephosphorylation of Tyr400 in virus budding process. The kinase inhibitors are found to show the defect in viral budding. Moreover, the kinase inhibitors are showed to inhibit the CP-cdE2 interactions at the surface of plasma membrane (Liu et al., 1996; Liu & Brown, 1993). Previous studies have suggested that cdE1 is not involved in interactions with CP (Barth et al., 1992). But the recent studies by Tang (Tang et al. 2011) provided the evidence for the involvement of E1 in virus budding. The first structure for the complete enveloped virus was demonstrated by cryo-electron microscopy and image reconstruction (Cheng et al., 1995). The high resolution cryo-EM structure of VEEV has been determined and its coordinates are available in public domain (Zhang et al., 2011). Combinations of cryo-EM and X-ray crystallographic studies have provided critical insights into the possible mode of interactions between the cytoplasmic tails of glycoproteins (E1 and E2) with CP (Zhang et al., 2011). Additionally, one more helix is observed in interdomain region of the capsid protein that might be involved in virus NC disassembly (Zhang et al., 2011).

The release of the virions from the surface of cellular membrane is known to be sensitive to ionic strengths. At the low ionic strength, the budding of virions become less efficient (Strauss et al., 1980; Waite et al., 1972). 6K protein is a viroporin that is involved in membrane permeability, glycoprotein trafficking and virus budding (Madan et al., 2008; Gonzalez & Carrasco, 2003). It also participates in the transport of pE2-E1 heterodimer at the location of virus assembly (Lusa et al., 1991).

1.10 Capsid protein structure

Alphavirus CP comprises of two domains, RNA binding N-terminal domain and C-terminal protease domain. The N-terminal domain is rich in basic amino acids and highly

disordered. The protease domain has a chymotrypsin like fold, possessing *cis*-autoproteolytic activity that splits itself from the nascent structural polyprotein. After the self-cleavage, it becomes inactive due to binding of C-terminus tryptophan to the active site (Skoging & Liljestrom, 1998; Choi et al., 1996; Tong et al., 1993; Choi et al., 1991). Till date, the crystal structures of CP from four members (SFV, SINV, AURAV and VEEV) of alphavirus family have been reported (Table 1.3).

Table 1.3: List of CP crystal structures from different members of alphavirus family

Alphavirus	PDB ID
Semliki Forest Virus (SFV)	1VCP, 1VCQ
Venezuelan Equine Encephalitis Virus (VEEV)	1EP5, 1EP6
Aura Virus (AURAV)	4AGK, 4AGJ, 4UON, 5G4B
Sindbis virus (SINV)	1WYK, 1SVP, 2SNW, 1KXA, 1KXB, 1KXC, 1KXD, 1KXE,

The cryo-EM and crystallographic studies of various alphaviruses provide key details about the possible mode of interactions of the various subunits of virus assembly. Structural analysis of protease domain of CP reveals two β -barrel sub-domains arranged in a Greek key motif (Fig. 1.10.1). The catalytic triad is present in between the two β -barrel sub-domains (Fig. 1.10.1). In all the serine proteases, including alphavirus CP, the catalytic triad residues and molecular architecture of active site is highly conserved.

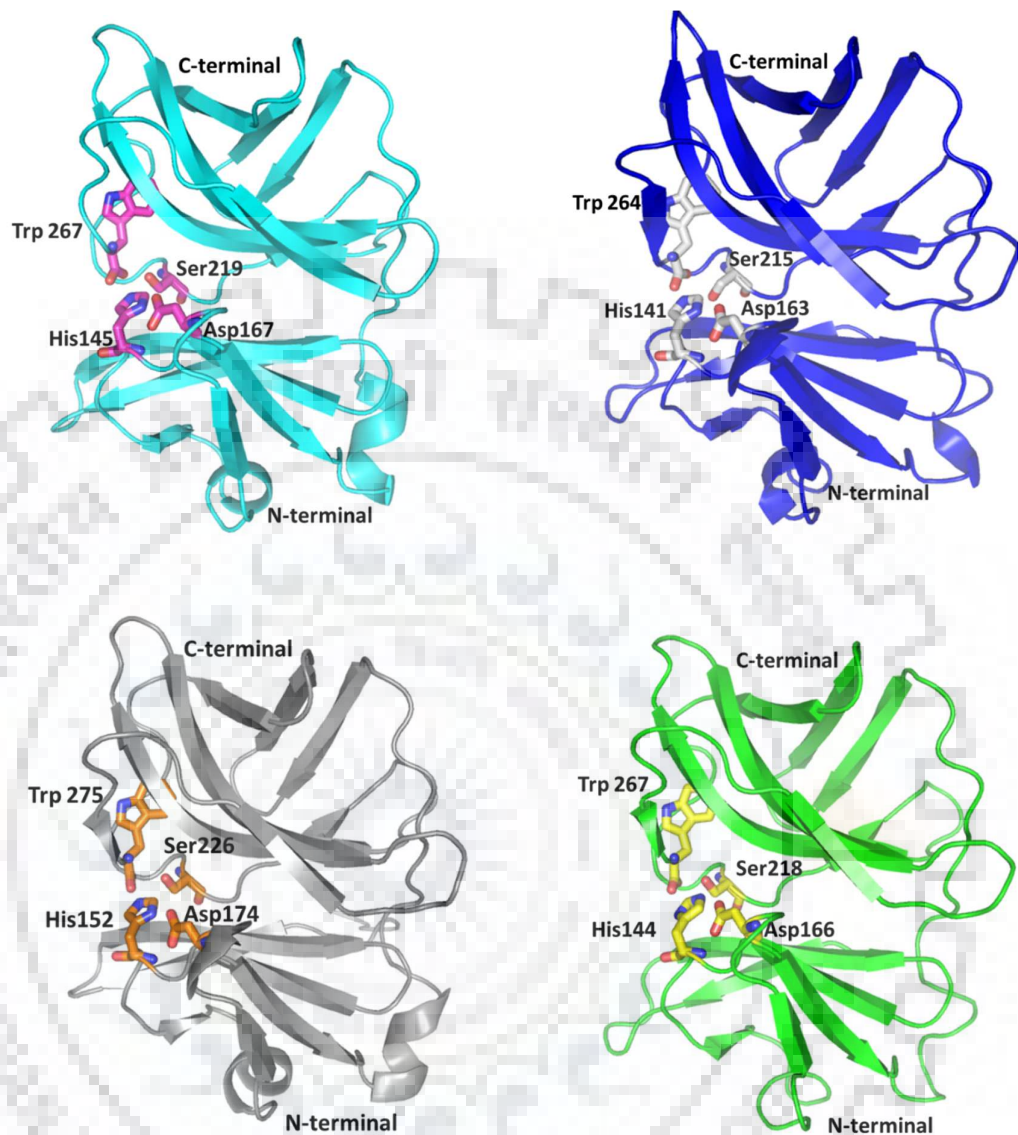


Figure 1.10.1. Cartoon view representation of alphavirus crystal structures. The cartoon view representation of crystal structure of SFCP (PDB ID: 1VCP), SCP (PDB ID: 1WYK), VEEVCP (PDB ID: 1EP5) and AVCP (4AGK) are shown in cyan, blue, grey and green color, respectively. The CP protease domain consists of two β -barrel domains with catalytic triad presents at the cleft between two subdomains (N and C terminal subdomains). The catalytic triad residues are presented as sticks. After the *cis*-proteolytic cleavage the C-terminal Trp remains bound to the active site and inhibit further protease activity. The terminal tryptophan residue bound to active site is also shown.

Similar to other serine proteases, the residues of catalytic triad in alphavirus CP are histidine, aspartate and serine. The mutation of the conserved serine residue of the catalytic triad, leads to loss of protease activity (Melancon & Garoff, 1987). Structural analysis of other chymotrypsin like proteases such as Rhinovirus 3C (HRV3C) and hepatitis A virus 3C (HAV3C) revealed the presence of cysteine residue at the active site rather than serine (Allaire et al., 1994;

Matthews et al., 1994). The CP protease domain is also involved in interactions with cdE2 glycoprotein and subsequently results in virus budding. The conserved hydrophobic pocket at the surface of the CP is a site for CP-cdE2 interactions. Therefore, it is suggested that the small molecules targeting the CP-cdE2 interactions could inhibit virus budding.

1.11 Alphavirus CP *trans*-activity

The CP protease domain has *cis*-proteolytic activity and in earlier literature stated not to possess *trans*-protease activity (Choi et al., 1991; Hahn et al., 1990). It cleaves itself at W/S scissile bond and after the cleavage, the terminal Trp interacts with the residues of active site and inhibits the further protease activity (Choi et al., 1991; 1997; Aggarwal et al., 2012; 2014). Hence, alphavirus capsid protein has proteolytic activity once in the virus life cycle (Hahn & Strauss, 1990). Conversely, the C-terminal truncated CP shows esterase activity (Morillas et al., 2008). The studies carried out by Morillas et al, showed that the truncation of seven residues from C-terminal SFCP results in restoration of CP protease activity.

The protease activity from many animal and parasitic proteases have been characterized in detail (Kumar et al., 2006; Yadav et al., 2013; Mohanty et al., 2003; Yadav et al., 2007). Protease inhibitors are attractive target for designing and development of new therapeutic agents (McKinlay, 2001; Anand et al., 2009). Till date, several FRET based proteolytic studies using the fluorogenic peptide substrate containing a FRET pair on both the sides have been observed (Mittoo et al., 2003; Angliker et al., 1995; Knight et al., 1992). The reduction of FRET signal has been observed after the cleavage of fluorogenic substrate peptide with active protease. FRET based protease assay has already been used for viral proteases like 3C proteases, HCV (hepatitis C virus) NS3/4A protease and SARS (Severe acute respiratory syndrome) coronavirus 3CL protease, Dengue NS3 protease (Blanchard et al., 2004; Balasubramanian et al., 2016; Sudo et al., 2005; Yang et al., 2013). The fluorogenic peptide substrate was derived from the sequence containing the protease specific scissile bond.

To find out the *trans*-protease activity of AVCP, last two residues were deleted from the C-terminal CP and FRET based protease assay was performed (Aggarwal et al., 2014). The purified AVCP truncated protein (AVCP Δ 2) was found to be fully active. The protease assay uses DABCYL (4-(4-dimethylaminophenyl-azo) benzoic acid) and EDANS (5-[(2-aminoethyl) amino] naphthalene-1-sulfonic acid) as FRET pair on the N and C terminus of CP cleavage site. In this FRET pair, EDANS is a fluorophore and DABCYL is a quencher, which quenches the fluorescence of EDANS placed at its certain proximity. The cleavage of fluorogenic peptide substrate generates two fragments separating DABCYL and EDANS which leads to reduction of

FRET signal and increase in fluorescence intensity. This method can be utilized for the testing of compounds libraries or FDA approved drugs against the viral specific protease activity and paves way for development of antiviral drugs against alphaviruses.



Chapter 2

Inhibition of chikungunya virus by picolinate that targets viral capsid protein

2.1 Abstract

This chapter studies the binding of picolinic acid (PCA) to the conserved hydrophobic pocket of capsid protein. The binding of PCA to chikungunya virus capsid protease (CVCP) hydrophobic pocket was analysed by molecular docking, isothermal titration calorimetry (ITC), surface plasmon resonance (SPR) and fluorescence spectroscopy. Additionally, PCA significantly inhibited CHIKV replication in infected Vero cells, decreasing viral mRNA and viral load as assessed by qRT-PCR and plaque reduction assay, respectively. It is postulated that the binding of picolinic acid to CVCP hydrophobic pocket might inhibit CP-cdE2 interactions seen in the cryo-EM structures of various alphaviruses (Kostyuchenko et al., 2011; Mancini et al., 2000; Sun et al., 2013; Zhang et al., 2011).

The alphaviruses capsid protein (CP) is a multifunctional protein. It is involved in autoproteolysis, nucleocapsid core formation, capsid-glycoprotein interactions, host transcription downregulation and contain signals for nuclear and cytoplasmic trafficking. Alphavirus CP, possesses chymotrypsin-like domain with conserved hydrophobic pocket at the surface, a site for capsid-glycoprotein interactions. The protein-protein interactions (PPIs) of the cytoplasmic domain of E2 glycoprotein (cdE2) with the hydrophobic pocket on the surface of capsid protein (CP) plays a vital role in alphavirus life cycle. The cryo-EM structure of VEEV unveiled that the loop region of a helix-loop-helix motif in the cdE2 fits into CP hydrophobic pocket. Moreover, the fitting analysis of crystal structure of AVCP-dioxane complex and cdE1, cdE2 models into VEEV cryo-EM density map, revealed the conserved Pro405 of AURAV cdE2 forms hydrophobic contacts with AVCP hydrophobic pocket. It was postulated that the dioxane bound to CP-hydrophobic pocket of AVCP-dioxane complex structurally imitates the pyrrolidine ring of Pro405 residue of cdE2. Previous studies have reported antiviral properties of PCA against Human Hepatitis B virus, Dengue virus and SINV. PCA was found to inhibit the SINV replicons at various concentrations tested. The chemistry of nitrogen containing heterocyclic rings can be exploited for design and synthesis of drug molecules. This study is suggestive of the potential of nitrogen containing heterocyclic rings as antivirals against alphaviruses.

2.2 Introduction

The re-emergence of chikungunya disease caused by CHIKV possesses a serious worldwide threat to the public health (Cavrini et al., 2009). The main vectors responsible for the transmission of CHIKV are *Aedes albopictus* and *Aedes aegypti*. CHIKV infection is characterized by acute fever, rashes and musculoskeletal pain in humans (Morrison, 2014). Apart from these symptoms, CHIKV fever is also characterized by persistent arthralgia (Larrieu et al., 2010). Clinical manifestations of CHIKV fever includes lethal hepatitis lymphopenia, severe dermatological lesions, encephalitis and fetal transmission during pregnancy causing neonatal encephalopathy (Sourisseau et al., 2007). The time between the mosquito bite to the start of the symptoms ranges from 2-12 days. This disease was first reported in 1952 from febrile patient serum following an outbreak on the Makonde plateau (Ross, 1956). This place is the border area near Mozambique and Tanzania. Since its first report in Africa, numerous outbreaks occurred in other parts of the world. In Kenya, major outbreak occurred in 2004 that had spread this disease in Comoro, Madagascar, Mayonette, La Reunion islands, South East Asia and West Africa (Sergon et al., 2008). An unexpected intrusion of the CHIKV infection in La Reunion Island in 2005 that affected one-third of the total population shocked the world (Nimmannitya et al., 1969; Padbidri and Gnaneswar, 1978; Sergon et al., 2007; Shah et al., 1964). In India itself, about more than 1.4 million people were affected with CHIKV epidemic in 2006 (Mavalankar et al., 2008). Prior to 2013, CHIKV infection had been reported in Asia, Africa, Europe, and the Indian and Pacific Oceans. Currently this disease has been identified in about more than 60 countries and territories (<http://www.cdc.gov/chikungunya/geo/index.html>). In 2008, US National Institute of Allergy and Infectious Diseases (NIAID) has listed this virus as category C priority pathogen (Thiboutot et al., 2010; Wang et al., 2008).

CHIKV is a member of genus Alphavirus that consists of enveloped single-stranded positive-sense RNA genome. These are transmitted by mosquitoes and infect wide range of animals including mammals. Alphavirus genus belongs to *Togaviridae* family and includes many medically relevant human and animal viruses including Ross River virus (RRV), Venezuelan Equine Encephalitis virus (VEEV), Chikungunya virus (CHIKV), Western Equine Encephalitis virus (WEEV) etc. The genus contains at least 29 members that are able to infect various vertebrates such as humans, horses, fish, birds and rodents. These members have been currently classified into eight complexes based on antigenic and genetic similarities (Powers et al., 2001; Weaver et al., 2012). Alphaviruses enter the host cells via receptor-mediated endocytosis followed by genome replication in the cytoplasm (Jose et al., 2009). The alphavirus virion is approximately 70 nm in diameter and has host cell derived lipid membrane which is embedded

with 80 spikes in T=4 icosahedral symmetry. Each spike is composed of trimers of E1 and E2 glycoprotein heterodimers (Mukhopadhyay et al., 2006; Soonsawad et al., 2010; Strauss and Strauss, 1994). The RNA genome (~11.7 Kb) is surrounded by nucleocapsid core (NC) that consists of 240 copies of CP. The virion genomic RNA with a 5' methylated cap and a polyadenylated tail is directly translated into viral replication enzymes in the host cell cytoplasm on entry (Baron et al., 1996; Vasiljeva et al., 2000). Alphavirus infection generally causes only an arthralgia, rash, headache, fever and arthritis. However, some of the pathogenic species like VEEV, EEEV, WEEV cause encephalitis and infections can also be life threatening. Alphaviruses are classified into New and Old World viruses depending upon the mechanism by which they down regulate the transcription of host cell and disease complications. Old World viruses like CHIKV, Semliki Forest virus (SFV) and SINV etc. employ nsp2 protein for down regulation of the cellular transcription and cause arthralgia. New World viruses like VEEV, WEEV etc. employ capsid protein for shutting off the host cell transcription and cause encephalitis (Garmashova et al., 2007; Hahn et al., 1988).

The CP of alphaviruses is a multifunctional protein. It consists of two domains, the RNA binding N-terminal domain and the C-terminal protease domain (Choi et al., 1991; Hong et al., 2006). The N-terminal domain, which is highly disordered is involved in binding to the genomic RNA, PPIs that lead to the CP dimerization and the host transcription shut off (Owen and Kuhn, 1996; Perera et al., 2001). Likewise, the alphavirus nsp4 protein has disordered N-terminal domain that involved in PPIs (Rupp et al., 2011; Tomar et al., 2006). The C-terminal domain of CP is a serine protease that possesses a chymotrypsin-like fold. It has autoproteolytic activity and splits itself from the nascent structural polyprotein and thus, plays a key role in the processing of structural polyprotein and the viral life cycle. After the *cis*-cleavage, the active site of the CP is occupied by terminal tryptophan residue that blocks further CP protease activity (Choi et al., 1991). Recently, *in-vitro* trans-protease activity of the C-terminal truncated AVCP and CHIKV CVCP using synthetic fluorogenic peptide containing the CP protease cleavage site has been reported (Aggarwal et al., 2014; Aggarwal et al., 2015).

The mature virion surrounded by two concentric shells consists of 240 copies of each of the capsid, E1 and E2 proteins. The outer shell comprises of the E1 and E2 glycoproteins and inner shell comprises of CP (Cheng et al., 1995; Strauss and Strauss, 1994). The space between the two concentric shells is filled with the host derived lipid bilayer through which the transmembrane helix of the glycoproteins penetrates (Mancini et al., 2000; Tang et al., 2011). The cryo-EM and crystallographic studies from various alphaviruses provide crucial details about the distribution, structures and possible interactions of the cytoplasmic tail of E1 and E2

glycoproteins and CP (Kostyuchenko et al., 2011; Zhang et al., 2002). The molecular interaction of CP with the cdE2 facilitates the budding of virus particles from the cellular membrane of infected host cell (Jose et al., 2012). To analyse this molecular interaction, the homology models of AVCP cdE2 glycoprotein has been generated and fitted into the cryo-EM density map of VEEV along with crystal structure of AVCP-dioxane complex. Fitting analysis revealed the hydrophobic contacts of the conserved Pro405 of AURAV cdE2 with the hydrophobic pocket of AVCP. The Pro405 is present at the loop region of helix-loop-helix structure of cdE2 (Aggarwal et al., 2012; Lopez et al., 1994; Owen and Kuhn, 1997). It was postulated that the dioxane present in the CP-hydrophobic pocket of AVCP-crystal structure structurally imitates the pyrrolidine ring of Pro405 residue of cdE2 (Aggarwal et al., 2012). This suggested that heterocyclic ring compounds similar to dioxane and its derivatives could bind to the CP hydrophobic pocket and disrupt its PPIs with cdE2. PCA, a pyridine containing compound is known to have antiviral, antimicrobial, cytotoxic and apoptotic properties (Corrêa et al., 2011; Fernandez-Pol et al., 2000). Previous studies have reported the antiviral properties of PCA against Human Hepatitis B virus, Dengue virus and SINV. PCA was found to inhibit the SINV replicons at various concentrations tested (Fernandez-Pol and Fernandez-Pol, 2008). The chemistry of nitrogen containing heterocyclic rings can be exploited for design and synthesis of drug molecules. Additionally, drug molecules based on pyridine ring such as isoniazid, the anti-tuberculosis agent are already available in the market (Preziosi, 2007). Recent findings have provided some breakthroughs in finding anti-CHIKV chemotherapeutic agents like Interferons, Ribavirin, Mercaptopurine etc. (Jadav et al., 2015). But till date, no effective drug or vaccine is available in the market for the treatment of CHIKV disease.

In this study, it is postulated that the binding of picolinic acid to CVCP hydrophobic pocket might inhibit CP-cdE2 interactions seen in cryo-EM alphavirus structures (Aggarwal et al., 2012; Tang et al., 2011; Zhang et al., 2011). Molecular docking, isothermal titration calorimetry (ITC), surface plasmon resonance (SPR) and fluorescence studies have been carried out to characterize the binding of PCA to the hydrophobic pocket of CVCP. Our results clearly demonstrate the antiviral properties of PCA against CHIKV, and thus support the postulate that the binding of PCA to the hydrophobic pocket of CVCP blocks cdE2-capsid interaction. Hence, this study illustrates the potential to use pyridine ring for design and development of pharmaceutically active antiviral compounds to combat chikungunya disease.

2.3 Materials and methods

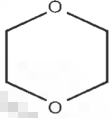
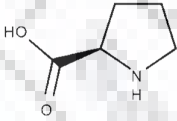
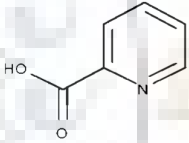
2.3.1 3D model generation

Homology models for CVCP protein (residue range: 113-261) were generated, in subsequent steps: (a) template selection from protein data bank (PDB), (b) sequence alignment, (c) model building, (d) refinement and validation. NCBI BLAST tool was used to carry-out the template search. The crystal structure of Semliki Forest virus capsid protein (PDBID: 1VCP) was the leading hit having 93 % sequence identity to CVCP (Choi et al., 1997). Clustal omega tool was utilized for the sequence alignment of the query with the template sequence (<http://www.ebi.ac.uk/Tools/msa/clustalo/>). 1VCP was used as a template to generate three dimensional structure of CVCP. Multiple models were made and the model with the lowermost DOPE score was further assessed using PROCHECK and ERRAT (<http://nihserver.mbi.ucla.edu/SAVES/>). The model with the minimum number of residues in the disallowed region was further subjected to energy minimization by means of Swiss-Pdb Viewer 4.01(<http://spdbv.vital-it.ch/>).

2.3.2. Molecular docking

AutoDock 4.2 was used for carrying out docking studies of small molecules into CVCP modelled structure. Docking was conducted using Windows 2007 on a HPxw8400 workstation. The modelled CVCP protein structure was selected for the docking studies. The protein structure was minimized and non- polar Hydrogens, Kollman charges (Kollman charges = 9) were added using AutoDock MGL Tools and save the protein in .pdbqt file format. The ligands for the CVCP protein docking studies were retrieved from PubChem Compound Database. The name of the ligands and their molecular formulas are given in Table 2.1. The ligands were downloaded from PubChem in SDF format and then converted in to PDB format using Open Babel software. For the ligand preparation, same procedure was employed for all the ligands. The Grid parameters were same. Hydrogens atoms and Gasteiger charges were added. For dioxane, proline and picolinic acid the Gasteiger charges added were -0.0002,-0.0001and -1.0001, respectively. After these changes the ligands were saved in .pdbqt file format. AutoGrid 4 was used to calculate the atomic potential grid map with a spacing of 0.375Å and the Grid box dimensions were 30 Å x 30 Å x 30 Å. The centre point co-ordinates set was X=8.726, Y=115.84 and Z=36.188. The number of total GA runs was increased from 10 to 50 and rest of the docking parameters were set as default. For interpretation, pose with the minimum binding energy was considered.

Table 2.1 List of the compounds with their empirical formula and molecular structure used for docking into the hydrophobic pocket of CVCP

Ligand	Molecular Formula	Molecular Structure
Dioxane	C ₄ H ₈ O ₂	
Proline	C ₅ H ₉ NO ₂	
Picolinic Acid (PCA)	C ₆ H ₅ NO ₂	

2.3.3 Virus

An Indian isolate of Chikungunya virus, DRDE-06 (GenBank accession no: EF-210157) belonging to East Central and South African (ECSA) genotype conserved at Virology Division, DRDE, Gwalior was employed in the present study for the expression of recombinant capsid (Aggarwal et al., 2015).

2.3.4 Expression and Purification of CVCP protein

2.3.4.1 Expression of CVCP protein

Chemically competent cells of *Escherichia coli* strain *Rosetta (DE3)* were transformed with recombinant plasmid containing N-terminal His-tagged CVCP capsid protein gene (residues 106-261). The amino acid residues for TEV protease cleavage site and the histidine affinity tag at the N-terminal of the CVCP protein was MGSSHHHHHSSSENLYFQGHM. The optimization of expression conditions for CVCP was done to produce the protein in soluble form. The LB agar plate was used for plating the transformed cells in presence of kanamycin (50 µg/ml)

and chloramphenicol (35 µg/ml), incubated at 37 °C overnight. One colony was picked and inoculated in LB broth supplemented with the desired antibiotics, grown at 37 °C overnight with constant stirring. The overnight grown culture was used to raise 1L secondary culture. The secondary culture was grown till the OD at 600 nm (OD₆₀₀) reaches to ~ 0.8 and afterwards the protein expression was induced by 0.4 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). After induction with IPTG, culture was grown at 37 °C for 4 h with constant agitation. Induced cells were harvested using centrifugation at 8,000 x g for 15 min at 4 °C and analyzed on 15 % SDS-PAGE (Sodium dodecyl sulfate-polyacrylamide gel electrophoresis) to confirm the protein expression and solubility.

2.3.4.2 Purification of CVCP protein

The CVCP protein was purified using cell pellet from 1L culture. Frozen pellets were thawed and re-suspended in 25 ml of lysis buffer (50 mM Tris-HCl pH 7.6, 10 mM imidazole and 100 mM NaCl) on ice. The re-suspended cells were disrupted using the French Press (Constant Systems Ltd, Daventry, UK) and centrifuged at 16,000 x g for 45 min at 4 °C to isolate the pellet and supernatant. The soluble protein was purified by Immobilized metal ion assisted chromatography (IMAC). The Ni-NTA (Nitrilotriacetic acid) agarose beads (Qiagen, USA) were pre-equilibrated with the purification binding buffer (50 mM Tris-HCl pH 7.6, 100 mM NaCl) and were loaded with the clarified supernatant in to the column. The supernatant was incubated with the Ni-NTA agarose beads at 4 °C for 45 min. Consecutive three washes were made with varied concentrations of Imidazole (W1=30 mM Imidazole, W2=50 mM Imidazole and W3=100 mM Imidazole) in combination with the binding buffer. Finally, the protein containing the N-terminal His-tag was eluted using 250 mM Imidazole. The eluted fractions were pooled and incubated with TEV protease for His-tag cleavage in dialysis buffer at 4 °C for 12 h. After the cleavage of affinity-tag, Ni-NTA column were refilled with protein samples to eliminate His-tagged TEV protease and uncleaved His-tagged protein. All the fractions were collected and analyzed using 15 % SDS-PAGE. For the efficient recovery of purified protein after dialysis, a dialysis membrane of 10 kDa MWCO was used (Thermo Scientific). The fraction containing His-tag cleaved protein was concentrated with an Amicon of 3 kDa cutoff (Millipore, USA). The UV absorbance spectroscopy at a wavelength of 280 nm was used for estimation of protein concentration using an extinction coefficient of 23,950 M⁻¹ cm⁻¹ for CVCP protein.

2.3.5 Circular dichroism (CD)

In the absence of the atomic structure of CVCP, the secondary structure of CVCP (residue range: 106-261) was determined using circular dichroism (CD) spectroscopy. The concentration of the protein used for CD measurements was 0.1 mg/ml for each protein sample. All the CD measurements were performed using PC controlled Applied Photophysics spectropolarimeter (Model Chirascan, UK). The spectra were recorded in a quartz cell of dimension 0.1 cm from 180 to 260 nm of wavelength with a bandwidth of 1 nm under continuous nitrogen purge. Three scans were recorded with a response time 0.25 s per point at 25 °C and afterwards the average of these scans was done. The baseline data was also collected for the buffer without protein and subtracted from the averaged spectrum of protein. The analysis of the data was done using the Dichroweb web server by CDSSTR method (Sreerama and Woody, 2000).

2.3.6 Isothermal titration calorimetry (ITC)

ITC experiments were conducted using MicroCal iTC₂₀₀ (GE Healthcare) to determine the thermodynamic parameters of dioxane and picolinic acid binding to CVCP protein. The titrations were executed at 25 °C in buffer containing 50 mM Tris (pH7.6), 20 mM NaCl. The picolinic acid used in the titration against CVCP was purchased from Sigma Aldrich (Cat. No. P42800). Picolinic acid (PCA) at the concentration of 12 mM (Syringe) was titrated in to CVCP protein 50 μM (Cell). Whereas the concentration of the dioxane used for the titrations was 1 mM (Syringe) against the CVCP protein 50μM (Cell). The various set parameters used to perform the isothermal titrations of dioxane and picolinic acid with purified CVCP are shown in Table 2.2. The ITC data obtained was subsequently analyzed by MicroCal Origin 7.0 software from MicroCal. The ITC data was integrated, dilution effects were corrected and concentration was normalized before the curve fitted to a one site binding model.

Table 2.2. The various set parameters used to perform the Isothermal titrations of dioxane and picolinic acid with purified CVCP

Parameters	Dioxane	Picolinic acid
Parameter	Setting	Setting
Total Number of Injections	30	30
Cell Temperature	25 ^o c	25 ^o c
Reference Power	8u ^{cal} /sec	8u ^{cal} /sec
Initial Delay	70sec	70sec
Syringe Concentration	1mM	12mM
Cell Concentration	0.05mM	0.05mM
Stirring Speed	1000 rpm	1000 rpm
Volume of 1st Injection	0.2ul	0.2ul
Duration of 1st Injection	0.4 sec	0.4 sec
Volume After 1st Injection	1.2 UL	1.4 sec
Duration After 1st Injection	17.13 sec	2.8 sec
Injection Spacing	280 sec	240 sec
Filter Period	5 sec	5 sec
Feedback Mode/Gain	High	High

2.3.7 Surface plasmon resonance (SPR)

BIACORE T200 (GE Healthcare, USA) instrument with CM5 sensor chips was used to perform all the SPR experiments. The analysis and sample compartment temperature was set 25 °C for all the binding and kinetic experiments.

2.3.7.1. Buffer Preparation

All the SPR binding and kinetic studies were carried out in Phosphate buffer saline (PBS). The PBS buffer, 10X (BR-1006-72) was purchased from GE-Healthcare. Distilled deionized water was filtered using 0.22 µm Millipore filters and degassed using Millipore Degassing unit to avoid microbubbles formation. As the analytes studied for binding and kinetic studies were small molecules, therefore 0.5% Tween20 (P9416-Molecular Biology grade) as a surfactant was added to the PBS 1Xbuffer. The final concentration of 1X PBS+P includes (20 mM NaH₂PO₄-Na₂HPO₄, 0.5% (v/v) surfactant Tween20 and 150 mM NaCl). This buffer was used for diluting the analyte stock solutions followed by preparation of samples and buffer blanks.

2.3.7.2 Immobilization of CVCP protein on CM5 sensor chip

The Amine Coupling Kit was purchased from GE Healthcare (GE Healthcare, BR-1000-50). The dextran surface of the CM5 chip was first activated with 1:1 mixture of 0.1 M N-

hydroxysuccinimide (NHS) and 0.1 M 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) with a flow rate of 10 μ l/min to form reactive succinimide esters. For immobilization of CVCP protein on surface of the sensor chip, the protein (100 μ g/ml) was dissolved in low salt buffer (0.1 M Sodium acetate buffer) with pH 4.5. After this treatment ligand acquired positive charge and got pre-concentrated onto negatively charged carboxymethyl dextran matrix of CM-Chip that further increase the efficiency of amine coupling. The CVCP protein was injected and flowed across the activated sensor surface with flow rate of 10 μ l/min. After the successful immobilization of ligand on the surface of chip, all the residual active succinimide esters were blocked using ethanolamine. The ligand (CVCP protein) was immobilized on flow cell 2 (FC-2) and the final immobilization level was 4,500 RU. For all the SPR experiments, the adjacent flow cell was used as a reference that contains no ligand and blocked in the similar way as activated flow channel. The stability of CVCP protein on the FC-2 surface was confirmed by five washes with regeneration buffer (10 mM Glycine-HCL, pH 2.5). The flow rate was kept constant throughout the binding (10 μ l/min) and kinetics experiment (45 μ l/min). The data was collected by Biacore control software and analysis was done with T200 evaluation software.

2.3.8 Fluorescence spectroscopy

Hitachi F-4600, fluorescence spectrophotometer was used to perform all the fluorescence experiments. The quartz cuvette of dimensions 10 mm x 4 mm was used for all the fluorescence measurements, recorded at 25 $^{\circ}$ C. To measure the intrinsic fluorescence of the CVCP, the protein sample (25 μ M) were excited at 280 nm and the emission spectra were recorded at wavelength of 295-550 nm. The slit width for excitation and emission was kept at 2.5 nm for recording the spectra. Each spectra obtained was average of three scans. The buffer comprising of 50 mM Tris, pH=7.6 and 20 mM NaCl was used for preparing the protein samples. The effect of picolinic acid (PCA) on the CVCP intrinsic fluorescence was studied. The emission spectra was recorded upon addition of PCA at different concentrations up to 5 mM (0.5, 1.0, 1.5, 2.0 and 5.0 mM) to the CVCP (25 μ M), after incubation for 15 min. The spectra obtained were compared with CVCP (protein without PCA) spectra and analysed for quenching.

2.3.9 Cells and virus propagation

The Vero cells (African green monkey kidney cell line) was procured from National Centre for Cell Science (NCCS), Pune, India and were maintained in minimum essential medium (MEM) supplemented with 10 % heat inactivated fetal bovine serum (FBS), 80 U gentamicin, 2 mM L-glutamine and 1.1 g/L sodium bicarbonate in humidified 37 $^{\circ}$ C, 5 % CO₂ incubator. An

Indian isolate of Chikungunya virus (strain DRDE-07; GenBank: EU372006) belonging to East Central and South African (ECSA) genotype maintained at Virology Division, DRDE, Gwalior was used for antiviral studies. Virus was propagated using standard virus adsorption techniques and titer was estimated by plaque assay (PFU/mL). Further, the virus stocks were stored in -80 °C for future use.

2.3.9.1 Cytotoxicity testing

The cytotoxic effect of the compound was determined by using neutral red dye uptake (NRDU) assay in Vero cells. The cells were seeded at cell density of 1×10^4 cells per well onto 96-well culture plates and incubated overnight for cell attachment. Serial two-fold dilution of compound was prepared in plates containing MEM from 8 mM to 62.5 μ M and incubated at 37 °C, 5 % CO₂ for 72 h. After incubation of 72 h, neutral red dye was supplemented into each well followed by incubation at 37 °C for one hour. The absorbance was recorded at 540 nm. The maximum nontoxic dose (MNTD) was determined. In situ Cell Death Detection Kit was used for carrying out TUNEL assay to detect apoptosis induced by PCA (16 mM - 2 mM) in Vero cells. The counterstained used was 4',6-diamidino-2-phenylindole (DAPI). This was further verified by Caspase 3 apoptosis assay, counterstained with Hoechst 33342 to identify nuclear morphology that allows comparison of apoptotic nuclei with surrounding normal nuclei that were observed microscopically at 36 hpi.

2.3.9.2 Antiviral assay

The seeding of Vero cells at cell density of 1×10^5 cells per well were carried out into 24-well culture plates and subsequent incubation overnight. The cells were infected with CHIKV at MOI of 0.01 excluding for the cell control. Compound was added 24 h before virus infection, 0 h and 2 h after the infection to reach a final concentration of 2 mM. The addition of compound 24 h before the infection is termed as pre-treatment. While the compound addition 0 h and 2 h after the infection is termed as simultaneous and post-treatment, respectively. The plates were incubated in 37 °C, 5 % CO₂ for 48 h. The infected supernatant was collected at 24 and 48 hpi for quantitation of both CHIKV RNA and infectious virus by qRT-PCR and plaque assay, respectively.

2.3.9.2.1 qRT-PCR assay

To determine the CHIKV load, Quantitative real-time RT-polymerase chain reaction (qRT-PCR) assay was performed. QIAamp viral RNA kit (Qiagen, Germany) was used for

isolation of CHIKV RNA from the supernatant and the isolated RNA was stored at -80 °C until further processing. An in-house qRT-PCR employing SS III Platinum one step qRT-PCR kit with primers differentiating CHIKV E1 segment was performed as described earlier (Agarwal et al., 2013). Standard curve for quantitation was prepared by qRT-PCR by utilizing sequential dilution of known copies number of *in vitro* transcribed RNA.

2.3.9.2.2 Plaque reduction assay

Plaque reduction assay was used to evaluate the compound's inhibitory effect on CHIKV infection. Vero cells were seeded into 24-well culture plates and incubated overnight. The 10-fold serial dilution of infected culture supernatant of compound treated-CHIKV cells along with virus control was prepared. After discarding medium, prepared dilutions of infected culture supernatant were used to infect the cell monolayer. The plate was incubated for virus adsorption for 1.5 h at 37 °C. The cell monolayer was further overlaid with medium comprising 1.25% methylcellulose followed by incubation for 3 days at 37 °C, 5 % CO₂. After the incubation, the overlaid medium was extracted and the plate was washed with PBS and fixed with chilled methanol before staining with crystal violet and the reduction of viral plaques compared to control was calculated.

2.4 Results

2.4.1 3D Homology model

Homology search of CVCP capsid protein gave twenty four hits against PDB database. Semliki Forest virus capsid protein (SFCP) crystal structure with 93% sequence identity (PDB - 1VCP) was the first hit. The homology models of CVCP capsid protein were generated using 1VCP as template. The cartoon representation of the CVCP 3D homology model was shown in Fig. 2.4.1.1. The Ramachandran plot analysis of the modeled CVCP 3D structure by PROCHECK, showed 97.3% and 2.7% residues are present in the favoured and allowed region, respectively. ProSA analysis (<https://prosa.services.came.sbg.ac.at/prosa.php>) revealed the folding energy of protein is well within the range of the ProSA plot that further validated the modeled structure. Fig. 2.4.1.1.

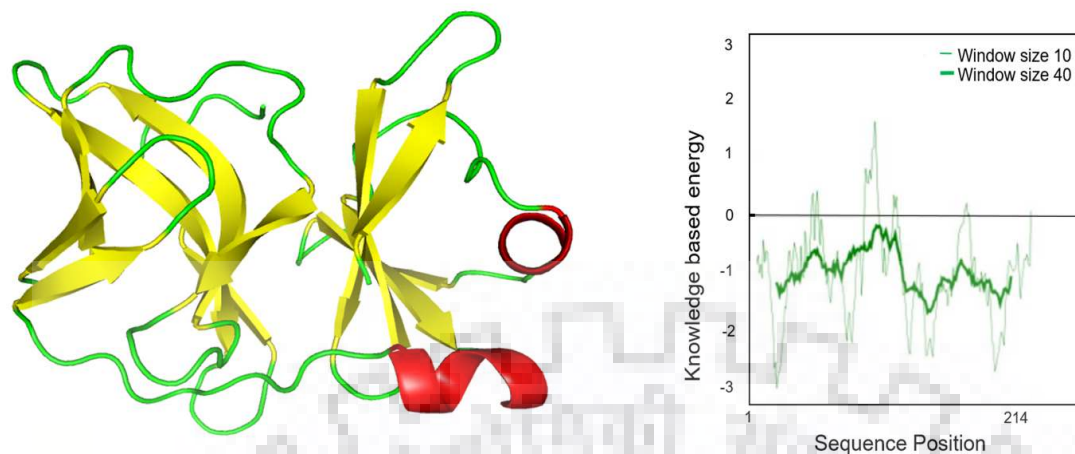


Figure 2.4.1.1: 3D atomic model structure of CVCP. Left panel shows cartoon diagram of CVCP representing helix (red colour), β - sheets (yellow colour) and loops (green colour). Right panel shows ProSA energy profile of CVCP modeled structure. The protein folding energy is in good agreement with the ProSA plot (Z-score -6.57).

The root mean square deviation (RMSD) of the CVCP model upon superimposition to the template 1VCP was 0.2 Å (Fig. 2.4.1.2).

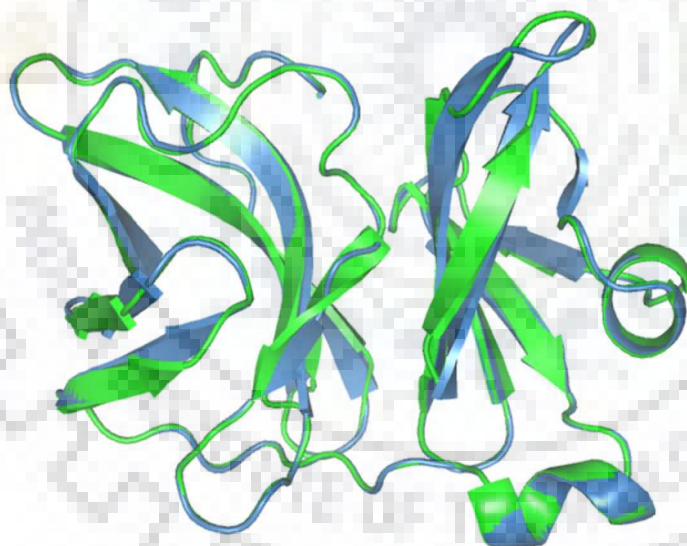


Figure 2.4.1.2: Structure comparison with alphavirus capsid protein. The cartoon representation of superimposition of CVCP homology model and SFCP (1VCP) template. CVCP and SFCP are shown in green and skyblue color, respectively.

As expected, the tertiary structure of CVCP exhibited similar geometry to that of crystal structure of SFCP (1VCP). The predicted structure of C-terminal domain of Chikungunya virus capsid protein (113-261) comprises of two β -barrel subdomains similar to crystal structure of other alphavirus CP's , reported till date. The subdomain I of CVCP (residue range: 113-261)

consists of five β -strands and two helices unlike the Semliki Forest virus capsid protein (residue range: 119-267) subdomain I that consists of six β -strands and same number of helices (Choi et al., 1997). Helix α -1 is present in between β -3 and β -4 strands and helix α -2 is present between β -5 and loop region connecting the two subdomains. The subdomain II consists of seven β -strands and devoid of any helix like other CP's. The two subdomains are connected by a linker loop region of 10 residues (Ala175-Glu184). The antiparallel strands of β -barrel subdomains [β 1 (114–119), β 2 (122–130), β 3 (133–137), β 4 (155–157), β 5 (161–166) of sub-domain I and β 6 (185–188), β 7 (193–197), β 8 (200–204), β 9 (214–218), β 10 (225–233), β 11 (237–245), β 12 (250–253) of sub-domain II] are folded in to two Greek keys, characteristic of chymotrypsin like serine protease. The capsid protein of alphavirus has *cis*-autoproteolytic activity that cleaves at scissile W/S bond and dissociates itself from the remaining structural polyprotein only once, in the virus life cycle. The further protease activity of CP was inhibited by C-terminal Trp residue as it blocks the active site of the protein (Tong et al., 1993). In between the two β -barrel subdomains the conserved catalytic triad, consisting of Ser213, His139 and Asp161 residues was present (Fig. 2.4.1.3). These catalytic residues architecture are conserved among the alphavirus CP's as well as other serine proteases. This protease domain of capsid protein cleaves the conserved scissile peptide bond between Trp-Ser residues. After the truncation of C-terminal Trp residue, CP protease activity in *trans* can be restored. Recent studies on FRET based assays have shown the *trans*-activity of C-terminal capsid protein (AVCP and CVCP) by using synthetic fluorogenic peptide containing cleavage site (Aggarwal et al., 2014; Aggarwal et al., 2015). Similar to the other chymotrypsin like serine proteases including alphaviruses, the GDSG motif containing the conserved active site nucleophilic serine residue is structurally conserved in CVCP (₂₁₁GDSG₂₁₄) (Choi et al., 1996; Forsell et al., 1995) (Fig. 2.4.1.3).

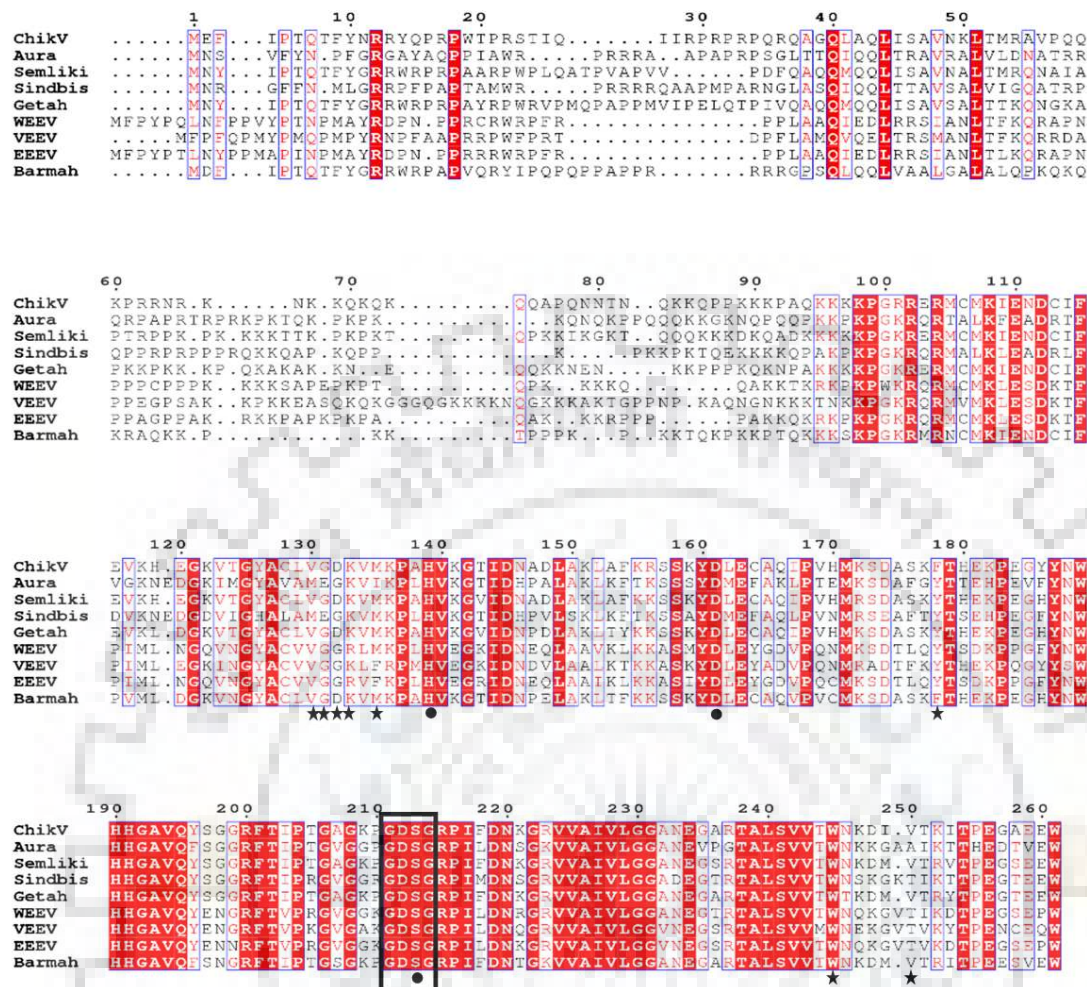


Figure. 2.4.1.3: Multiple sequence alignment of CVCP protein with sequences of other alphavirus capsid protein (CP). The conserved residues are highlighted in red background. The star under the amino acids represented the hydrophobic pocket residues of CP's involved in interaction with Dioxane, Proline and Picolinic acid. The highly conserved catalytic residues (His139, Asp161 and Ser213 in CVCP) of CP's are represented by solid circles below the residues. The GDSG motif ($_{211}$ GDSG $_{214}$) containing the conserved active site nucleophilic serine residue is shown in the box.

2.4.2. Molecular docking

Docking studies were performed using dioxane, proline and PCA as ligands and their binding in to the hydrophobic pocket of CVCP was analysed. Previous studies suggested the role of dioxane and its derivatives in disrupting the capsid-cdE2 interactions and thus alphaviral replication (Kim et al., 2007; Kim et al., 2005). Recent published data suggested the proline residue (Pro 405 in AVCP) of helix-loop-helix (H-L-H motif) of cdE2 fits into the hydrophobic pocket of CP and is involved in CP-cdE2 PPIs (Aggarwal et al., 2012). The crystal structures of CP-dioxane complex demonstrate dioxane molecule binding in the CP hydrophobic pocket that mimics the pyrrolidine ring of cdE2 proline. Thus, docking of proline and its derivative PCA into the hydrophobic pocket of CVCP was done to analyze ligand binding. The binding energy

of proline (-3.63 kcal/mol) and picolinic acid (-3.23 kcal/mol) into the CVCP hydrophobic pocket was found to be more than that of dioxane (-2.86 kcal/mol) (Table 2.3). The interactions of dioxane, proline and picolinic acid with the hydrophobic residues of the CVCP were analysed and shown in Fig. 2.4.2.1. Picolinic acid was found to interact with CVCP hydrophobic pocket residues through 4 hydrogen bonds. While, dioxane and proline bound to hydrophobic pocket through 2 Hydrogen bonds each. This is suggestive of that PCA and its derivatives might be more effective in preventing CP-cdE2 PPI and thus inhibit the alphavirus replication

Table 2.3. The values of binding energy obtained from molecular docking of dioxane, proline and picolinic acid into the CVCP hydrophobic pocket

CVCP Modelled Structure	DIOXANE	PROLINE	PICOLINIC ACID
Binding Energy (kcal/mol)	-2.86	-3.63	-3.23

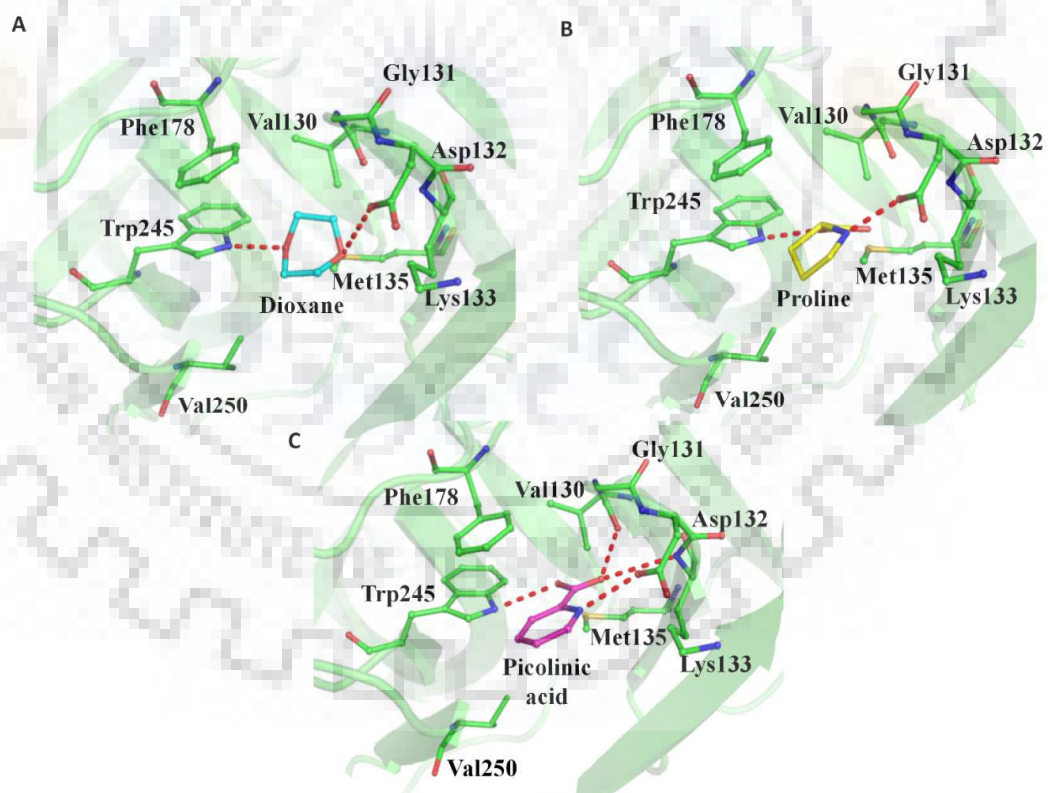


Figure. 2.4.2.1: Binding analysis of docked compound into CVCP hydrophobic pocket. The diagram showing the binding of (A) dioxane, (B) proline and (C) picolinic acid in the hydrophobic pocket of CVCP. The residues of the hydrophobic pocket involved in hydrophobic interactions are shown in green color and dioxane, proline and picolinic acid are shown in cyan, yellow and magenta color, respectively. The polar interactions of dioxane (2 hydrogen bonds), proline (2 hydrogen bonds) and picolinic acid (4

hydrogen bonds) with the hydrophobic pocket residues are shown in red dashes. Picolinic acid binds to CVCP hydrophobic pocket more strongly than dioxane and proline.

2.4.3 Purification of CVCP and secondary structure analysis

The soluble fraction of CVCP was purified using Ni²⁺ affinity chromatography and SDS-PAGE (15 %) analysis showed a single protein band of ~16 kDa. The bands of purified protein with and without His-tagged TEV protease are shown in Fig. 2.4.3.1.

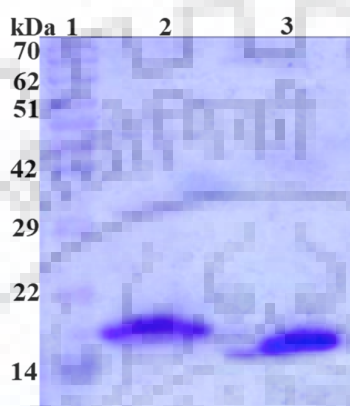


Figure 2.4.3.1: SDS-PAGE analysis of chikungunya virus capsid protein (CVCP). The SDS bands show the protein purified to homogeneity. Lane 1, molecular-weight markers (kDa); Lane 2, purified CVCP lane with His-tag (uncleaved); Lane 3, purified CVCP lane without His-tag (cleaved).

The TEV protease was used to remove N-terminal 6xHis-tag from the purified CVCP. The secondary structure of the CVCP was determined using circular dichroism analysis. The online web server Dichroweb was used for the data analysis. The deconvoluted data reveal the secondary structure content of CVCP was 5 % α -helix, 36 % β -sheet, 25 % β -turn and 32 % random coil. CD results show that purified CVCP used in this study has well defined secondary structures similar to other alphavirus capsid protease and is natively folded.

2.4.4 Isothermal studies of picolinic acid binding

The binding of dioxane and picolinic acid to the purified CVCP was studied using MicroCal iTC₂₀₀ at 25 °C by following the heat effect after injection of these ligands into cell containing protein. The ITC binding isotherms were fitted as single set of site model by using Origin7 Data Analysis software (GE Healthcare) and shown in Fig. 2.4.4.1. The binding constant of dioxane and PCA were calculated. The thermodynamic analysis of dioxane and picolinic acid binding to CVCP is shown in Table 2.4. There was stronger affinity of CVCP protein for picolinic acid with a K_D of 61 μ M than dioxane with a K_D of 80 μ M.

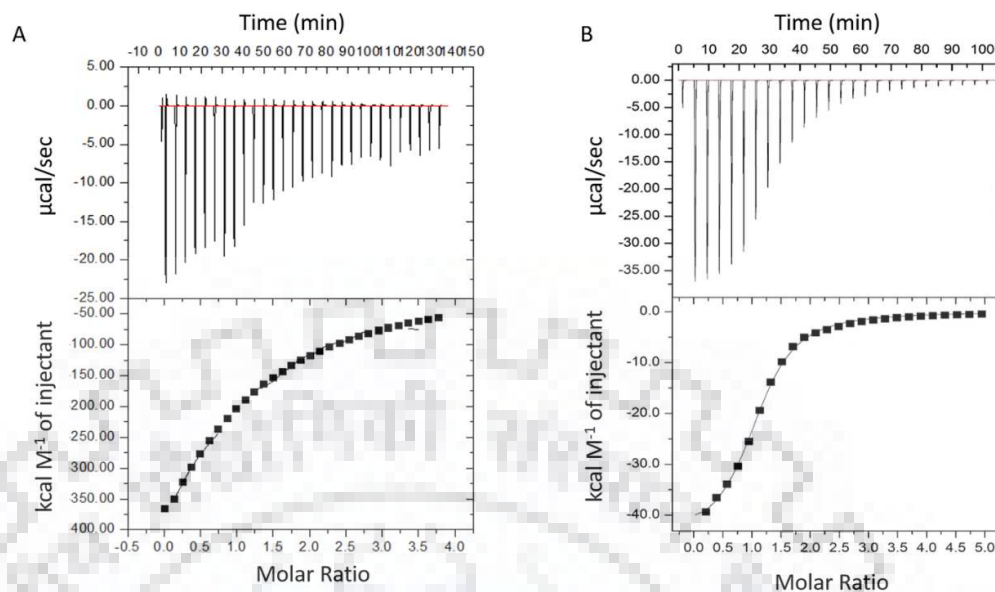


Figure. 2.4.4.1: ITC raw and integrated data. (A) Isothermal Titration Calorimetry (ITC) raw and integrated data for CVCP-dioxane titrations. The titrations were performed at 25 °C in buffer containing 50 mM Tris (pH7.6), 20 mM NaCl. Dioxane at the concentration of 1 mM (Syringe) was titrated in to CVCP protein 50 µM (Cell). (B) ITC raw and integrated data for CVCP-picolinic acid titrations. The temperature and the buffer components were similar to that of CVCP- Dioxane titrations. Picolinic acid at the concentration of 12 mM (Syringe) was titrated in to CVCP protein 50 µM (Cell). The experimental data points for the titrations were represented in solid squares and best fit of these data points were shown as solid curve by taking one site binding model.

Table 2.4. The thermodynamic analysis of dioxane and picolinic acid binding to CVCP as obtained from ITC

Protein	Ligand	N	K_D (µM)	K_A (M^{-1})	ΔH (kcal/mol)	ΔS (cal/mol/degree)
CVCP	Dioxane	1	80	$(12.6 \pm 0.9)10^3$	-1040 ± 0.8	-3.4×10^3
CVCP	Picolinic acid	1	61	$(16.3 \pm 0.9)10^3$	-44.6 ± 0.6	-130

2.4.5 Surface plasmon resonance

To get the quality sensor data the cleaning of instrument with DESORB followed by preconditioning of sensor chip was done. Standard cleaning procedure was performed using 0.5% (w/v) SDS and 50mM Glycine pH=9.5 as the running buffer and then primed for five to six times using distilled deionized filtered /degassed water. The SPR instrument was then kept in the standby mode overnight. Next day, a fresh CM5 sensor chip (research-grade) was docked

into the clean instrument followed by preconditioning hydrates and cleans the CM5 layer (carboxymethyl dextran).

2.4.5.1 CVCP protein immobilization

We used standard amine-coupling method for the immobilization of CVCP protein. The final immobilization level achieved was 4,500 RU on the surface of Flow cell 2 (FC-2). Flow cell 1 (FC-1) was selected as a reference to minimize the unwanted drift and systematic noise. The desired immobilization level was achieved by choosing inbuilt “aim for immobilization” method of Biacore control software. The specified “aim for immobilization” for CVCP protein was 5,000 RU and the final immobilized level was 4,500 RU. In this method a sequence of short pulses of the CVCP protein were injected over the activated surface until the required immobilization level was attained. After the immobilization step, the surfaces of both the reference (FC-1) and activated (FC-2) flow channels of the sensor chip was blocked by injecting 1.0 M ethanolamine (pH 8.5) with the flow rate of 20 $\mu\text{l}/\text{min}$. PCA was injected over the immobilized CVCP, at concentrations ranging from 0.0625 mM to 2 mM, for binding studies. The responses obtained upon binding, from these varied analyte concentration range was used to measure the kinetic parameters of the interaction. Each PCA concentration was tested in duplicate that defines the activity of immobilized CVCP is retained throughout the experiment. Reproducibility of the results rules out any mishandling and further supports the regeneration buffer used was appropriate for disrupting PCA/CVCP complex after each cycle without affecting the CVCP activity. For kinetic studies the contact time of PCA binding to the CVCP protein was monitored for 120 s with the flow rate of 30 $\mu\text{l}/\text{min}$ which was sufficient to get a nice binding response. To demonstrate the stability of the base line during the experiment, blank injections (only running buffer) were injected along with the PCA injections. The responses obtained for the buffer injections were flat, indicating the stability of the baseline.

2.4.5.2 Kinetic analysis

Kinetic parameters for binding of PCA to the immobilized CVCP protein were measured. The data obtained was evaluated by the Biacore Evaluation software (Biacore T200 Evaluation Software, Version: 2.0). The data was fitted as per the global fitting of the kinetic model with one state model and stoichiometry 1:1 (Fig. 2.4.5.1). Final experiment reports showed no bulk factor which ensures the proper sample preparations and handling. The values of the dissociation constants (K_D) obtained from SPR for binding of PCA to CVCP were comparable to the values obtained from ITC studies. The K_D value obtained for binding of PCA to CVCP from ITC and

SPR was 61 μM and 0.21 μM , respectively. This difference in dissociation rate constants could be due to steric hindrance caused by immobilization strategy, kind of buffers used in performing the experiments, set parameters and principle behind these two biophysical techniques (ITC and SPR).

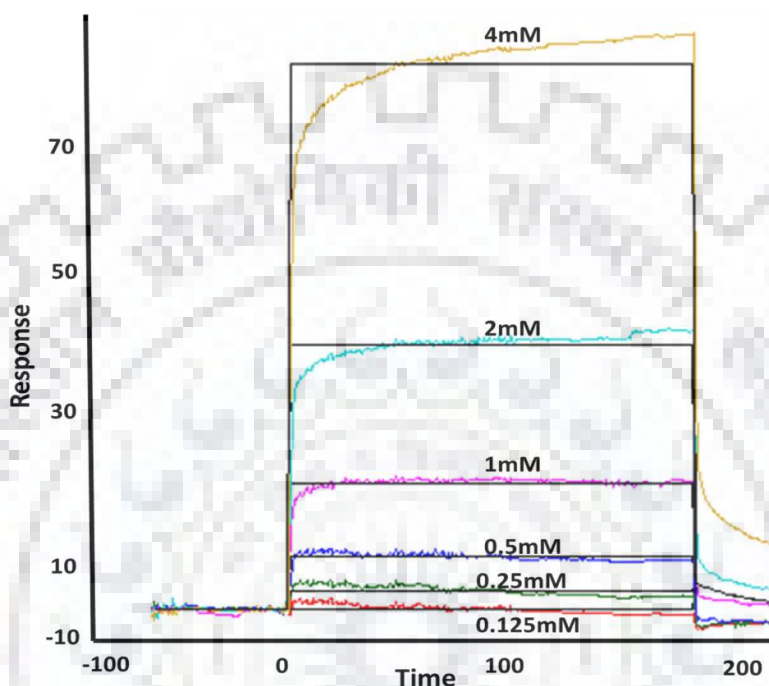


Figure. 2.4.5.1: SPR binding and kinetic analysis. The relative responses obtained for the binding of analyte (PCA) to the immobilized CVCP protein. The data was fitted by 1:1 binding state model. The K_D value obtained for the binding of PCA to immobilized CVCP was 0.21 μM .

2.4.6 Effect of PCA on CVCP intrinsic fluorescence

The intrinsic fluorescence of CVCP was measured in its native form (residue range: 106-261) and upon binding to PCA. As suggested from docking studies of PCA with CVCP, the PCA binds to the hydrophobic pocket of CVCP, rather more strongly than previously reported dioxane. The binding of the PCA to CVCP hydrophobic pocket results in decrease of fluorescence intensity with the increase in concentration of PCA as shown in Fig. 2.4.6.1. The binding constant for PCA binding to CVCP estimated from ITC and SPR was comparable. Fluorescence studies further validated the ability of PCA to bind to CVCP hydrophobic pocket that is expected to disrupt the CP-cdE2 PPIs.

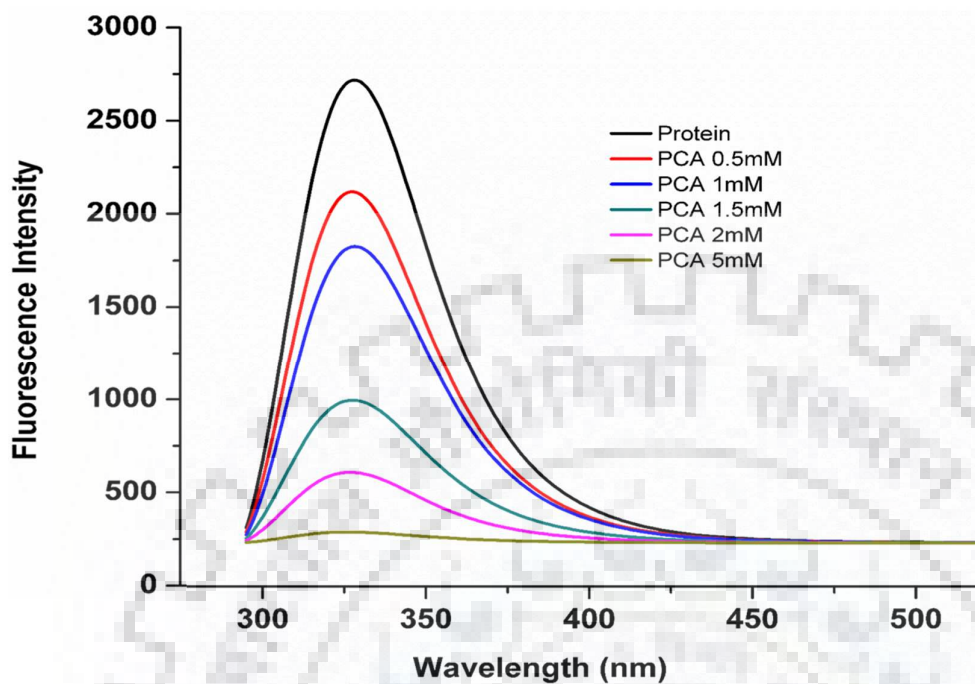


Figure. 2.4.6.1: Fluorescence spectroscopy. The intrinsic fluorescence of CVCP native protein and upon binding to PCA was determined. The emission spectra was recorded upon addition of PCA at different concentrations up to 5 mM (0.5 mM, 1.0 mM, 1.5 mM, 2.0 mM and 5.0 mM) to the native CVCP (25 μ M), after incubation for 15 min. The PCA was supposed to bind to CVCP hydrophobic pocket and thereby decreased the fluorescence intensity with the increase in concentration of PCA

2.4.7 Cytotoxicity testing

To determine the cytotoxic effect of this compound before subsequent assays, the maximal nontoxic dose (MNTD) was determined on Vero cells by the conventional neutral red dye uptake assay. The MNTD value of the compound in this study was found to be approximately 2 mM that showed more than 90 % cell viability. Both TUNEL and Caspase 3/FITC assay followed by DAPI and Hoechst 33342 counterstaining, respectively of PCA treated Vero cells did not show apoptosis at 2 mM PCA. Thus, 2 mM PCA concentration was used to carry out all the subsequent experiments (Fig. 2.4.7.1).

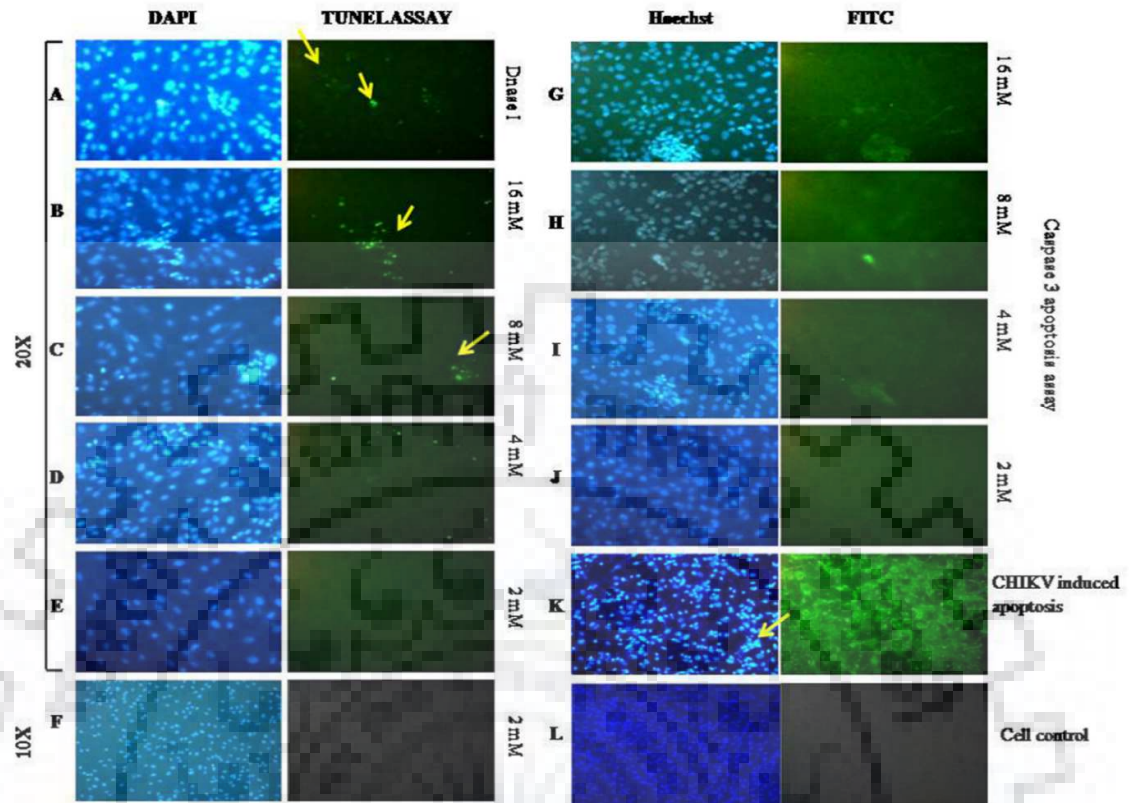


Figure 2.4.7.1: Apoptosis in PCA treated Vero cells. Micrographs of (A-F) TUNEL/DAPI-stained and (G-J) Caspase3-FITC/Hoechst-stained, PCA treated Vero cells. (K) CHIKV treated Vero cells were taken as positive control, showing densed chromatin/nuclear fragmentation within the nuclei of cells undergoing apoptosis (see arrows) and (L) cell control 10X. Negative control without TdT enzyme from TdT reaction buffer was also kept aside (data not shown).

2.4.8 Antiviral assay

The antiviral activity of the compound against CHIKV was determined by using qRT-PCR. Viral RNA was isolated from the infected culture supernatant of compound treated and untreated cells, and was quantitated. The inhibition of CHIKV RNA load as depicted in Fig. 2.4.8.1 was significantly reduced by up to 2 log in simultaneous treatment ($p < 0.01$). The treatment of infected cells with compound showed highly significant inhibition in simultaneous-treatment up to 48 h compared to virus control. Marginal inhibition was observed in pre-treatment up to 24 h and no inhibition was observed in post-treatment.

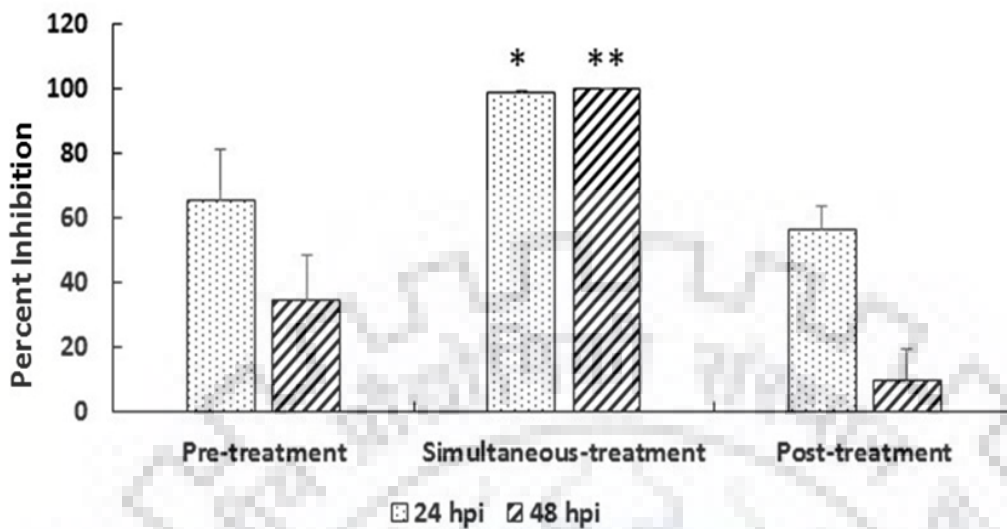


Figure 2.4.8.1: qRT-PCR. Viral RNA copy number by qRT-PCR. Vero cells were treated with PCA in pre-treatment, simultaneous treatment and post-treatment modes (infected with 0.01 MOI of CHIKV). Cell supernatants were harvested at 24 and 48 hpi and RNA was isolated. The qRT-PCR was performed with specific primers for CHIKV E1 region. Titer of CHIKV RNA was determined from a standard curve drawn using 10-fold serially diluted *in vitro* transcribed CHIKV RNA.

The antiviral activity of the compound against CHIKV was also confirmed by conventional plaque assay. Our results indicated that the CHIKV load was considerably reduced following compound treatment by 60.63 % in simultaneous treatment at 48 hpi ($p < 0.01$). Approximately 45 % reduction was observed in pre-treatment only up to 24 hpi. However, no significant reduction was seen in post-treatment study (Fig. 2.4.8.2).

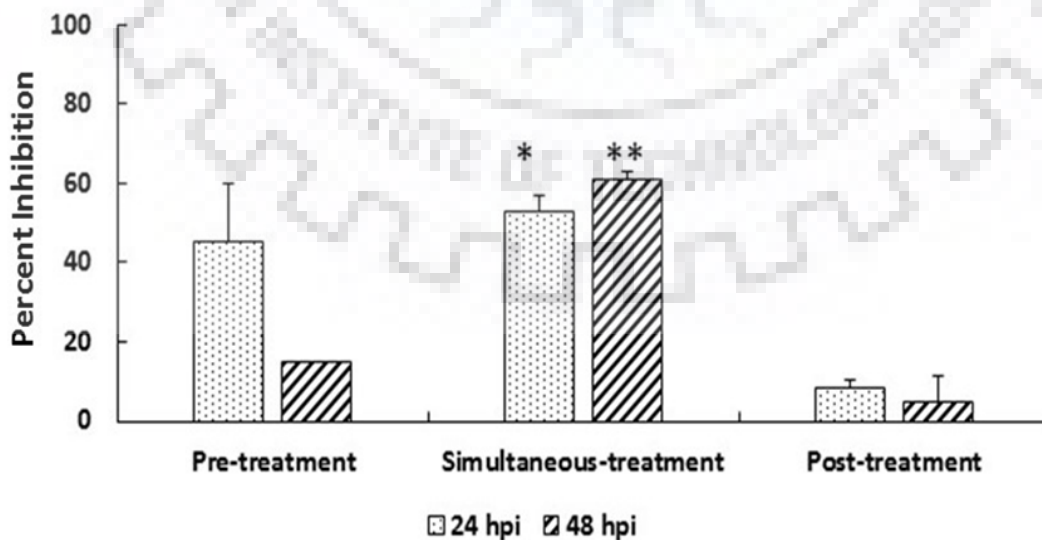


Figure 2.4.8.2: Plaque reduction assay. Relative CHIKV titer as assessed by plaque assay at 24 and 48 hpi. Percent inhibition with respect to virus control was calculated. Data represents the mean \pm standard deviation of three independent experiments. The asterisk indicates statistical significance (* p <0.05, ** p <0.01).

2.5 Discussion

The genus *Alphavirus* includes notable human and animal pathogens. The members of this genus divided in to two main groups: Old and New World alphaviruses. They have different disease presentation and different mechanism to shut the host transcription off. Old World viruses such as SINV, SFV, CHIKV etc. are arthralgic in disease presentation and employ nsp2 protein for shutting the host transcription off. New World viruses such as EEEV, VEEV etc. are encephalitic in disease presentation and employ capsid protein for inhibition of host transcription (Garmashova et al., 2007; Hahn et al., 1988). The N-terminal domain or RNA binding domain of capsid protein involve in this mechanism of transcription inhibition and not the protease domain. However, over the last few decades, chikungunya disease associated with neurological and haematological complications have been reported. Recent outbreaks of chikungunya, from 2005 onwards have been reported to be associated with encephalitis and meningoencephalitis (Chandak et al., 2009; Chusri et al., 2011). The re-emergence of chikungunya as an epidemic, evolution of new mosquito vector of the virus (due to mutation in E1 glycoprotein) (Schuffenecker et al., 2006) and the pathogenicity of virus especially encephalitis and meningoencephalitis apart from arthralgia makes it a significant pathogen which needs proper and immediate attention. Till date, there is no vaccine to prevent chikungunya disease and no specific medicine to treat it. This necessitates, the study of alphaviral biology for specific anti-CHIKV drug development.

In the current study, we focused on finding inhibitors that bind the hydrophobic pocket of CP, and obstruct the CP and cdE2 glycoprotein interactions. The interaction of cdE2 with nucleocapsid core is critical for the virus life cycle. Although the crystal structures of CP from other alphavirus members like SINV, AURAV, SFV and VEEV have been reported, the crystal structure of Chikungunya virus capsid protein is yet to be determined. The cryo-EM studies of CP and cytoplasmic domain of E2 glycoprotein (cdE2) interactions provide important breakthrough in understanding this key interaction. Previous studies have found a solvent-derived dioxane molecule bound to hydrophobic pocket in the crystal structure of AVCP, SCP (Aggarwal et al., 2012; Choi et al., 1991) suggesting that dioxane and its derivatives may prevent the capsid-cdE2 interactions. Many Dioxane derivatives were extensively studied in Sindbis virus and found to have antiviral property (Kim et al., 2007; Kim et al., 2005). Alphavirus capsid protein is a serine protease with a chymotrypsin-like fold and a conserved hydrophobic pocket

at the surface for interaction with transmembrane glycoproteins. The crystal structure of alphavirus capsid protein (AVCP, SCP, and SFCP) has structurally conserved hydrophobic pocket on the CP surface. Therefore, ligands that bind the CP hydrophobic pocket more strongly than dioxane are expected to inhibit capsid–cE2 interaction in various alphaviruses.

The CP possesses cis-autoproteolytic activity and it cleaves itself from the rest of the structural polyprotein only once, in the virus life cycle. The subsequent protease activity of CP was inhibited by C-terminal Trp residue (AVCP, W-267) as it blocks the P1 position near the active site. After the truncation of C-terminal Trp residue, CP protease activity in trans can be restored (Aggarwal et al., 2014). Like other serine protease, CHIKV CP has a catalytic triad consists of His139, Ser213 and Asp161 residues. The catalytic triad is positioned between the two β -barrel sub-domains. Due to the absence of crystal structure of Chikungunya CP (CVCP), homology models were generated by modeller 9.14. After homology modelling the 3D structure obtained was validated and energy minimized. AutoDock 4.2, was used for studying the docking of dioxane and PCA into the hydrophobic pocket of CVCP. The binding energy of picolinic acid (-3.23) into the CVCP hydrophobic pocket was found to be more than that of dioxane (-2.86). It clearly reflects that PCA binds to the CP hydrophobic pocket more strongly and efficiently than the previously reported dioxane. The thermodynamic parameters of the binding were calculated based on the isothermal titration calorimetry and surface plasmon resonance. The binding constant (K_D) calculated for the binding of dioxane and PCA (picolinic acid) to CVCP hydrophobic pocket from ITC was 80 μ M and 61 μ M respectively. This result was further validated by SPR. The K_D value of PCA binding to CVCP hydrophobic pocket obtained from SPR kinetic experiment was 0.21 μ M. These results of PCA binding indicate that other small molecules related to PCA and its derivatives that bind to the alphavirus hydrophobic pocket could block the capsid assembly and thus inhibit the viral replication as well as capsid-cE2 interactions.

Prior to the estimation of antiviral properties of PCA, its cytotoxic effect has been determined. Cell viability assay revealed that MNTD value of the PCA, was approximately 2 mM in Vero cells that showed more than 90% cell viability at 72 h. Further, this concentration (2 mM) of PCA did not show apoptosis as determined using TUNEL and Caspase 3 apoptosis assay. In the present investigation, we determined the inhibitory effects of PCA against CHIKV in the pre, simultaneous and post-treatment in Vero cells. The computational modelling studies indicated that PCA binds to the hydrophobic region of CHIKV capsid protein, which was also confirmed by fluorescence spectroscopy and SPR results. Thus, hypothesized that PCA should have antiviral activity against CHIKV via inhibiting the disassembly and assembly functions of

the CP or the virus budding process. Our results showed that pre-treatment of cells with PCA had partial effect on CHIKV replication that supports the hypothesis that PCA is inhibiting the initial disassembly. Whereas, in the simultaneous treatment, PCA inhibited CHIKV up to 48 hpi by affecting virus fusion and assembly process. Interestingly, no antiviral effects were observed in the post-treatment, indicating that PCA does not interfere with the process of virus release, confirming that the main inhibitory effect of PCA is at the step of disassembly, replication and/or assembly of nucleocapsids. However, it is observed that the post-treatment after 24 hpi, but not after 48 hpi, inhibited viral RNA abundance. This may be representing the inhibition of re-infection by newly released virions at the early disassembly step. Interestingly, at 48 hpi no inhibition at RNA level is observed and this could be due to overabundance of capsid protein. Therefore, at this stage enough amount of PCA is not available to saturate the hydrophobic pockets of all the CP molecules and thus, not able to inhibit CHIKV. These results clearly demonstrate the antiviral activity of PCA against CHIKV, supporting the postulate that PCA binds to the hydrophobic pocket of CVCP and inhibits its function. However, the exact mechanism by which PCA inhibits CHIKV replication will be elucidated by performing additional *in vitro* and *in vivo* capsid assembly and disassembly experiments. In nut shell PCA at concentration 2 mM significantly inhibited CHIKV replication in simultaneous-treatment mode, decreasing viral mRNA and viral load as assessed by quantitative real-time RT-PCR and plaque reduction assay respectively, in infected Vero cells. Hence, this study paves the way for developing more precise and efficient antivirals against re-emerging CHIKV and other pathogenic alphaviruses. Of particular note, outstanding inhibitory effects against alphaviruses may be observed following treatment with PCA in combination with other antiviral drugs. However, further *in vivo* experimental studies are required to investigate their potential application for clinical intervention of alphaviruses.

2.6 Conclusions

Molecular docking results showed stronger binding of the PCA (-3.23 kcal/mol) to the hydrophobic pocket than that of the dioxane (-2.86 kcal/mol). The binding constant (K_D) calculated using isothermal titration calorimetry (ITC) for dioxane and PCA were 80 μ M and 61 μ M respectively confirming the results of molecular docking. The binding kinetics of PCA with CVCP were further studied using surface plasmon resonance (SPR). The dissociation constant K_D (M) obtained for PCA was 0.21 μ M. The intrinsic fluorescence spectroscopy confirms the binding of PCA in the hydrophobic pocket. Additionally, PCA was found to have inhibitory activity against CHIKV replication in Vero cells, decreasing viral mRNA and viral

load as assessed by qRT-PCR and plaque reduction assay, respectively. This study suggests antiviral potential of heterocyclic ring compounds similar to PCA and its derivatives against CHIKV and alphaviruses.



Chapter 3

Structure-function insights into chikungunya virus capsid protein: Small molecules targeting capsid hydrophobic pocket

3.1 Abstract

The present chapter reports the crystal structure of chikungunya virus capsid protein (CVCP) from the diffraction data extending up to 2.2 Å resolution. The CVCP structure is compared to the capsid protein structures from other members of alphavirus family. Like other alphaviruses CP's, CVCP has serine protease activity with hydrophobic pocket at the surface, a site for capsid-glycoprotein interactions. The protein-protein interactions (PPIs) between CP hydrophobic pocket and cytoplasmic tail of E2 (cdE2) are essential for alphavirus budding. CHIKV cdE1 and cdE2 models were generated utilizing the VEEV cryo-EM structures as search template. To analyse the CP-cdE2 interaction, CVCP crystal structure and the cdE1, cdE2 models were fitted in to VEEV cryo-EM density map. The structural analysis unveils new insights in to CP-cdE2 interactions. The CHIKV cdE2 suggested to have helix-loop-helix (H-L-H) conformation through which it interacts with the surrounding residues of the CVCP hydrophobic pocket. The conserved Tyr400 residue of cdE2 exhibit the hydrophobic contacts with Leu411 and Leu412 of another helix (helix-II) of helix-loop-helix motif of cdE2. Moreover, the conserved Cys396 residue of helix-I of helix-loop helix motif forms disulfide bond with Cys417 lying at the loop region at the end of helix-II and provides the stability to the H-L-H motif. The sequence comparison of CVCP with other alphavirus genus has also been performed for the studies of different regions of protein.

The CVCP protease domain has been expressed in bacterial expression system and purified by affinity column chromatography using Ni-NTA beads and gel-filtration chromatography. The concentrated CVCP protein was used for crystallization and the crystals were obtained by vapor diffusion method by using 0.1 M BisTris (pH 6.5), 0.1 M Octyl β-D-glucopyranoside (BOG), and 20 % w/v polyethylene glycol (PEG) 3,350 at 20 °C. The obtained rectangular shaped crystal was diffracted at home source at 2.2 Å. The space group for CVCP crystal was triclinic (*P*1) and the parameters for unit cell were $a=33.3$, $b=39.1$, $c=59.3$ Å. Final CVCP model has estimated Matthews coefficient of 2.29 (Å³ Da⁻¹) and contains two molecules per asymmetric unit.

3.2 Introduction

Chikungunya fever is an arthropod-borne, re-emerging viral disease, reported recently in Central and South America, apart from Asia and Africa (Abdelnabi et al., 2017; Cavrini et al., 2009; Morrison, 2014). Chikungunya virus (CHIKV) has been accountable for causing recent epidemics of chikungunya disease in the countries of Indian subcontinent and many other parts of the world. The main vectors responsible for CHIKV transmission are *Aedes aegypti* and *Aedes albopictus*. In humans, CHIKV infection is associated with a range of complications from mild fever, musculoskeletal pain, rashes to persistent arthralgia. Clinical manifestations of CHIKV disease include lymphopenia, severe dermatological lesions, encephalitis and fetal transmission during pregnancy causes neonatal encephalopathy (Larrieu et al., 2010; Sourisseau et al., 2007). The disease was first reported in 1953, in Makonde plateau from the serum of a febrile patient, a place at the border near Mozambique and Tanzania (Ross, 1956). Since its first incidence, numerous outbreaks occurred in various parts of the world. In 2004, major outbreak occurred in Kenya that had spread the disease in Mayonette, Comoros, Madagascar, La Reunion Island, South East Asia, Europe and West Africa (Sergon et al., 2008; Sergon et al., 2007). In 2005, CHIKV outbreak in La Reunion Island had shocked the world by affecting one third of its population (Gerardin et al., 2008; Soumahoro et al.). In India itself, CHIKV epidemic in 2006, had affected more than 1.4 million people (Kumar and Gopal, 2010; Mavalankar et al., 2008; Padbidri and Gnaneswar, 1978; Shah et al., 1964). CHIKV has been listed as category C pathogen by US National Institute of Allergy and Infectious diseases (NIAID) in 2008 (Schwartz and Albert, 2010). Unfortunately, till date, no antiviral drug or vaccine is commercially available against CHIKV infection.

CHIKV is a member of genus Alphavirus and family Togaviridae. Alphaviruses are group of enveloped, single-stranded positive sense RNA viruses carried by arthropod vectors. Other members of this genus include many human and animal viruses like Ross River Virus (RRV), Semiliki Forest Virus (SFV), Western Equine Encephalitis virus (WEEV), Sindbis Virus (SINV), Venezuelan Equine Encephalitis virus (VEEV) etc. These are further classified into Old World viruses and New World viruses based on the mechanism employed for shutting the host transcription off, mortality rate and disease presentation. Old World viruses include CHIKV, SINV and SFV etc. utilize the nsp2 protein to down regulate the host cell transcription, have low mortality rate and cause arthralgia. New world viruses include VEEV, WEEV etc. utilize their capsid protein to down regulate host cell transcription, have high mortality rate and are encephalitic in disease presentation (Garmashova et al., 2007; Hahn et al., 1988). The genus alphavirus comprises of 29 members, able to infect a range of animal hosts such as humans,

rodents, fish, horse etc. Additionally, based on antigenic and genetic similarities these are further categorized into eight complexes out of which CHIKV belongs to the Semiliki Forest Virus antigenic complex (Jadav et al., 2015; Powers et al., 2001; Weaver et al., 2012). Like other alphaviruses, CHIKV has genome of approximately ~11.7 kb, capped at 5' end and polyadenylated at 3' end. The RNA genome is encapsidated by around 240 copies of capsid protein (CP), forming nucleocapsid core (NC) (Baron et al., 1996; Vasiljeva et al., 2000). The mature alphavirus virion (70 nm in diameter) has envelop derived from host cell membrane, embedded with 80 spikes in T=4 icosahedral symmetry. Each spike is made up of trimers of E1 and E2 (surface glycoproteins) heterodimers (Mukhopadhyay et al., 2006; Soonsawad et al., 2010; Strauss and Strauss, 1994).

Alphavirus CP is a multi-functional protein consisting of two domains, namely RNA binding N-terminal domain and C-terminal protease domain (Choi et al., 1991; Melancon and Garoff, 1987; Strauss and Strauss, 1994). The N-terminal domain being less conserved and highly disordered with high degree positive charge, is responsible for binding to RNA genome, shutting host transcription off and dimerization of CP (Lulla et al., 2013; Owen and Kuhn, 1996). Likewise, alphavirus N-terminal nsp4 protease is highly disordered and believed to involve in protein-protein interactions (PPIs) (Shirako et al., 2000; Tomar et al., 2006). The C-terminal domain of alphavirus CP is highly conserved and is a chymotrypsin-like serine protease. It possesses cis-proteolytic activity that cleaves at the W/S scissile bond and separates CP from the structural polyprotein (Choi et al., 1991; Melancon and Garoff, 1987). After the cis-proteolytic cleavage, the C-terminal Trp residue of CP left bound to the S1 specificity pocket of the protease and blocks the protease activity (Aggarwal et al., 2012; Choi et al., 1991). Recently in vitro trans-protease activity using fluorogenic peptide containing CP protease site have been reported for Aura virus capsid protein (AVCP) and CVCP (Aggarwal et al., 2014; Aggarwal et al., 2015). Crystal structure studies have revealed that the basic molecular architecture of the active site and the catalytic triad are highly conserved among the serine proteases including the CP of alphaviruses (His139, Asp161 and Ser213 CVCP residues) (Aggarwal et al., 2012; Choi et al., 1991). Furthermore, the GDSG motif that contains the active site serine nucleophile is also conserved among alphavirus CPs (211GDSG214 in CVCP) (Aggarwal et al., 2012; Choi et al., 1996; Choi et al., 1991). Alphavirus CP (1-261) is ~29 kDa protein and contains nuclear localization signal (NLS) and nuclear export signal (NES) for nuclear-cytoplasmic trafficking. In CHIKV, the CP amino acid stretches from residues 60-99 and 143-155 have been mapped as NLS and NES, respectively (Thomas et al., 2013). *In silico* studies have shown that the binding mode of CHIKV CP to exportin proteins through NES is similar to that of snurportin-1 binding

to exportin proteins (Thomas et al., 2013). Moreover, the CP of chikungunya virus contains linear B-cell epitopes that have been identified and mapped to the N-terminal (residues 1-35) and the C-terminal (residues 157-175) domains of CHIKV CP (Goh et al., 2015; Kam et al., 2014).

Molecular interactions between the cdE2 and the hydrophobic pocket present in the C-terminal CP domain play crucial role in the virus budding process. Besides molecular contacts between cdE2-CP, the cdE1-CP interaction has also been reported to be important for the budding process (Barth et al., 1992). Combinations of X-ray crystallographic and cryo-electron microscopy (cryo-EM) structural studies have provided crucial insights into the possible mode of interactions of E1 and E2 glycoproteins with CP (Zhang et al., 2011). Structural analysis of CP and glycoprotein structures fitted into the cryo-EM density map of alphavirus revealed that the conserved Pro405 residue of Aura virus (AURAV) cdE2 makes molecular contacts with the conserved CP hydrophobic pocket (Aggarwal et al., 2012). AURAV cdE2 has helix-loop-helix (H-L-H) conformation with Pro405 in the loop region. The bound dioxane in the AVCP hydrophobic pocket, structurally imitates the pyrrolidine ring of Pro405 residue of cdE2 (Aggarwal et al., 2012; Lopez et al., 1994; Owen and Kuhn, 1997). Moreover, recent findings have demonstrated that the binding of picolinic acid (PCA) to the hydrophobic pocket of CVCP inhibits chikungunya virus replication (Sharma et al., 2016). This suggested that heterocyclic ring compounds similar to dioxane and PCA have potential to bind the hydrophobic pocket and disrupt its PPIs with viral glycoproteins.

This study aims to structurally characterize CVCP, investigate the CVCP-cdE2 glycoprotein interactions and to identify inhibitors that bind into the conserved CP hydrophobic pocket and potentially block the viral budding process. In this work, the crystal structure of CVCP has been determined at 2.2 Å resolution. The crystal structure of CVCP and the structure models of glycoproteins were fitted into the VEEV cryo-EM density map to analyse CP-cdE2 interactions. Based on the molecular interaction, small drug-like heterocyclic molecules were docked into the CVCP hydrophobic pocket. The binding of molecules was experimentally evaluated using surface plasmon resonance and fluorescence spectroscopy. Additionally, CVCP structure was used for identification and analysis of linear B-cell epitopes in the CP of CHIKV. The availability of CVCP structure paves way for design and development of inhibitors based on identified epitopes and small heterocyclic drug-like molecules targeting the conserved CP hydrophobic pocket of CHIKV.

3.3 Materials and methods

3.3.1 Expression and Purification

The methods for expression and purification of CVCP have been described previously (Sharma et al., 2016). Concisely, the gene containing C-terminal domain (residues 106-261) gene was cloned in pET28c vector with an N-terminal TEV protease cleavable His-tag. The CVCP protein was overexpressed in *E. coli* Rosetta (DE3) strain using isopropyl b-D-1-thiogalactopyranoside (IPTG) and was purified by Ni-NTA affinity chromatography. Further, TEV protease was added to remove the His-tag and purified protein fraction was used for gel-filtration chromatography by using HiLoad 16/60 prep grade Superdex 75 column (GE Healthcare) (Fig. 3.3.1.1). CVCP purified protein was concentrated and subjected to crystallization trials.

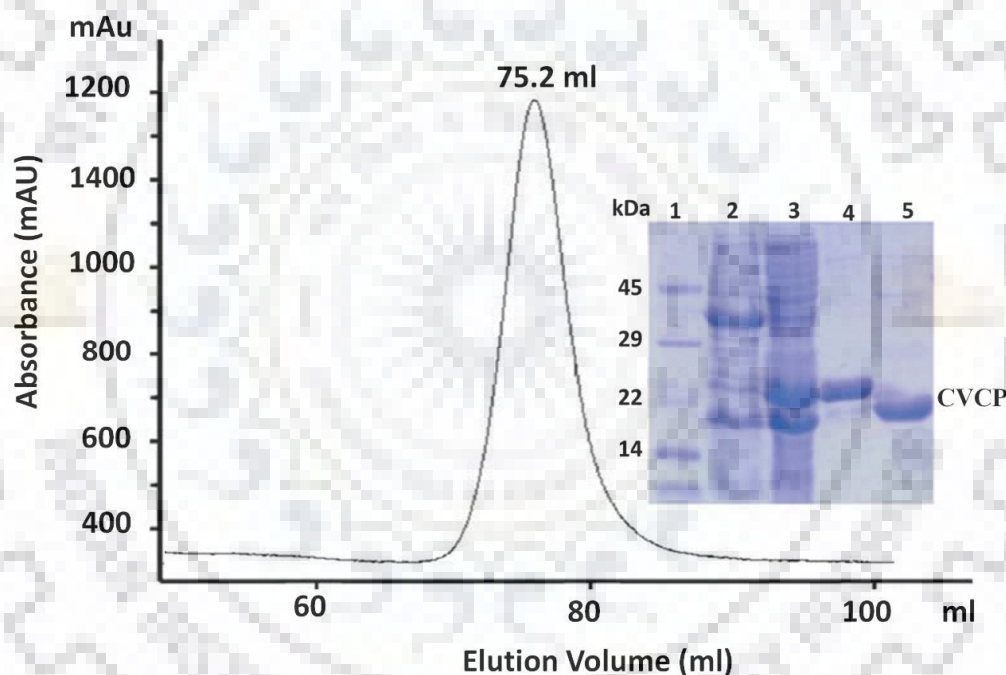


Figure 3.3.1.1: Gel filtration chromatography and SDS-PAGE analysis of chikungunya virus capsid protein (CVCP). Gel filtration shows the monomeric nature of protein. The SDS bands show the protein purified to homogeneity. Lane 1, molecular-weight markers (kDa); Lane 2, Pellet; Lane 3, Supernatant; Lane 4 Purified CVCP lane with His-tag (uncleaved) and Lane5, Purified CVCP lane without His-tag (cleaved).

3.3.2 Crystallization and data collection

The sitting drop vapor diffusion technique was used for protein crystallization. The purified CVCP protein was concentrated in 3kDa cut off amicon (Millipore) to ~8 mg/ml and subjected for crystallization in buffer containing 50 mM Tris-HCl (pH 7.6.), 5 % glycerol and 20 mM NaCl. The ratio of protein and reservoir buffer used for setting up crystallization trays

was 1:1 against the 100 μ l reservoir buffer. Crystals were obtained at 20 $^{\circ}$ C in 100 μ M BisTris (pH 6.5), 20 % w/v polyethylene glycol (PEG) 3,350 and 100 μ M octyl β -D-glucopyranoside (BOG); the latter was used as an additive for improving crystal quality. Use of BOG during crystallization significantly decreased the mosaicity and improved the size of the CVCP crystals. The initial hits grew gradually and full size crystals grew over the course of 4 weeks. The obtained rectangular shaped crystal of CVCP was subjected to diffraction at home source (Fig. 3.3.2.1). The cryoprotectant composition was also optimized for improving the resolution. The crystal was equilibrated for 2 min in cryoprotectant solution composed of mother liquor containing 25 % (v/v) ethylene glycol. Data was collected at the home source with MAR 345 imaging plate system using Cu K α radiation generated by a Bruker-NoniusMicrostar H rotating-anode generator operated at 45 kV and 60 mA (Fig. 3.3.2.1). The data was collected at cryogenic conditions (100 K) at a wavelength of 1.54 \AA . The diffraction data was processed using HKL2000 package (Otwinowski and Minor, 1997). The data collection and refinement details are summarized in Table 1.

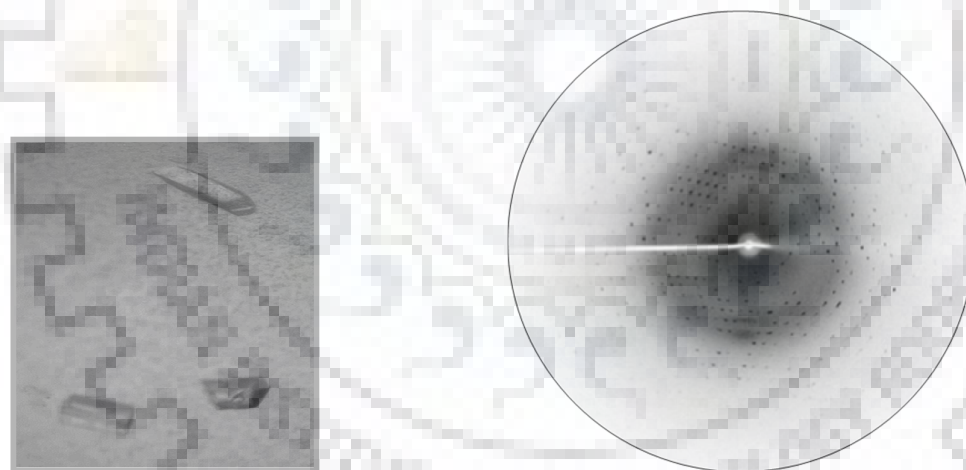


Figure 3.3.2.1: Crystal image (left) of chikungunya virus capsid protease (106-261). Diffraction image (right) of CVCP crystals using in-house radiation source at the Macromolecular Crystallographic Facility (MCU), IIC. The resolution obtained at the edge of the plate is 2.2 \AA .

3.3.3 Structure solution and refinement

The crystal structure of CVCP was determined using molecular replacement method and Semliki Forest virus capsid protein (SFCP) (PDB ID: 1VCP) crystal structure was employed as the search model. The MOLREP program was used for carrying out molecular replacement (Vagin and Teplyakov, 1997), and the PHENIX 6.0 (Adams et al., 2010) was used for refinement.

The electron density map analysis and manual model building were carried out using the COOT program (Emsley and Cowtan, 2004). Model rebuilding was done by examining Sigma-A-weighted maps. Structural analysis of the refined model and the preparation of figures were done using the PyMOL visualization tool (The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger) (DeLano, 2002). The stereo-chemical properties of the refined models was evaluated by using MOLPROBITY (Davis et al., 2007, Jain et al., 2010). Surface assembly analysis was carried out by PISA (Shoemaker et al., 2009).

PDB accession number: The PDB accession number for deposited CVCP structure coordinates is 5H23.

3.3.4 Molecular modeling

Chikungunya virus E1 and E2 glycoprotein models were made using the program MODELLER and Swiss-Model (Sali and Blundell, 1993; Schwede et al., 2003). The template used for building CHIKV E1 and E2 models was structure of VEEV (PDB ID: 3J0C) (Zhang et al., 2011). Clustal omega program was used for multiple sequence alignment of query with template sequence (<http://www.ebi.ac.uk/Tools/msa/clustalo/>). The quality of the models were evaluated by PROCHECK, ProSA plot and SAVES server (Wiederstein and Sippl, 2007). The model with the least number of residues in the disallowed region was selected and energy minimization was performed using Swiss-Pdb Viewer 4.01 (<http://spdbv.vital-it.ch/>). The E1, E2 models and the crystal structure of CVCP was fitted into VEEV cryo-EM density map (EMDB ID: 5275) by using Chimera fit-in-map module (Pettersen et al., 2004). Though the EM structures often have low resolution, but the combination of X-ray structures and their fitting into cryo-EM density map provide insights of viral subunits localization at an atomic level.

3.3.5 Identification and docking of small heterocyclic molecules

The small heterocyclic molecules structurally similar to PCA were identified from PubChem database. AutoDock 4.2 was used for carrying out docking of identified small molecules into CVCP hydrophobic pocket. Docking was conducted using Windows 2007 on a HPxw8400 workstation. The 3D structures of ligands for the CVCP protein docking were retrieved from PubChem Compound Database. The name of the ligands and their molecular formulas are given in Table 2. Hydrogens atoms and Gasteiger charges were added. For dioxane, picolinic acid, MDA and EAB the Gasteiger charges added were 1.9874, 0.9877, 0.9876, and 1.9875, respectively. AutoGrid 4 was used for calculation of atomic potential grid maps with

0.375 Å spacing and 34 Å x 34 Å x 34 Å grid box dimensions. Obtained results were manually analyzed using PyMOL (DeLano, 2002).

3.3.6 Surface plasmon resonance

BIACORE T200 instrument (GE Healthcare, USA) was used for executing all the SPR experiments. SPR binding and kinetic studies were carried out in Phosphate buffer saline [10 mM phosphate buffer with 2.7 mM KCl (pH 7.4), 137 mM NaCl, 0.5 % DMSO]. Additionally, 0.1% Tween20 (P9416-Molecular Biology grade, Sigma) was added in the PBS buffer to reduce non-specific binding and for suitable solubility of analytes. This buffer was used for preparation of test samples from the stock solutions of analytes and also as a control buffer sample for performing SPR experiments. The refractive index change at constant flow rate over a period of time was recorded. The net increase in the refractive index as compared to the control recorded over time, corresponds to the amount of ligand that binds to CVCP. This net change is recorded in response unit (RU). The data was collected using Biacore control software and analysis was done using Biacore evaluation software.

Standard amine-coupling method was used for the immobilization of CVCP protein. CVCP (100 µg/ml) dissolved in 100 µM Sodium acetate buffer (pH 4.5) was injected over CM5 sensor surface for carrying out immobilization. After the activation of flow cell surface by flowing 1:1 mixture of 100 mM N-hydroxysuccinimide (NHS) and 100 mM N-ethyl-N'-(dimethylaminopropyl)-carbodiimide (EDC), the CVCP protein was injected. CVCP protein was injected to the activated flow cell surface at a flow rate of 10 µl/min for 7 min. The adjacent flow cell was used as a reference for doing all the SPR measurements that contains no ligand and blocked in the similar way as activated flow channel.

3.3.7 Fluorescence spectroscopy

Intrinsic fluorescence experiments for purified CVCP protein were performed using a Hitachi F-4600 fluorescence spectrophotometer. The quartz cuvette of dimensions 10mm x 4mm was used for recording the fluorescence spectra, at 25 °C. To measure the fluorescence intensity of CVCP, the protein sample (10 µM) were excited at 280 nm and the emission spectra were recorded in the range of 295-550 nm wavelength. Each spectra obtained was the average of three scans. The buffer containing 50 mM Tris, pH 7.6 and 20 mM NaCl was used for diluting the stock solutions of protein and ligand molecules. The effect of MDA and EAB on the CVCP intrinsic fluorescence was studied. The emission spectra was recorded upon addition of MDA (2.5, 5.0, 7.5 and 10 mM) and EAB (6.25, 25, 50, 100, 200 and 400 µM) to CVCP (10 µM), after

incubation for 15 min. The spectra obtained were compared with native CVCP spectra and analyzed for quenching.

3.3.8 *In-silico* epitope identification

The *in silico* tools were utilized to identify B-cell epitopes present on the surface of CHIKV capsid protease domain. The linear B-cell epitopes were predicted by ABCpred (Saha and Raghava, 2007) and Bepipred (Larsen et al., 2006). The antigenicity was evaluated using Kolaskar and Tongaonkar antigenicity scale (Kolaskar and Tongaonkar, 1990). These methods used the properties of amino acids and their rate of occurrence in experimentally identified epitopes. The biochemical properties like surface accessibility was evaluated using Emini surface accessibility scale (Emini et al., 1985), the flexibility of protein segments was evaluated using Karplus and Schulz (Karplus and Schulz, 1985) and hydrophilicity was predicted using Parker Hydrophilicity Prediction method (Parker et al., 1986).

3.4 Results and Discussion

3.4.1 Structure determination and overall architecture

CVCP protein was crystallized in the triclinic space group (*P1*) with two molecules per asymmetric unit. The estimated Matthews coefficient was 2.29 ($\text{\AA}^3 \text{ Da}^{-1}$) which correspond to 49.0 % solvent content. The structure of CVCP was determined from diffraction data extending to 2.2 \AA resolution. BLASTp results showed that CVCP sequence shares highest (~93%) sequence homology with sequence of SFCP (PDB ID 1VCP). The molecular replacement method was employed and the crystal structure of SFCP was utilized as structure solution model. The Ramachandran plot of the final model showed 100% residues lying in the allowed regions. The final refined model covers 2510 non-hydrogen protein atoms (region 111 to 261), 88 solvent molecules and 63 non-hydrogen atoms corresponding to ligands (Octyl β -D-glucopyranoside (BOG), glycerol and ethylene glycol). In CVCP crystal structure, the density of first five residues (106-110) was missing from the N-terminus. BOG is confined near the N-terminus of both CVCP monomers and suggesting that it provides stabilization to the N-terminus of CVCP. The model was refined with PHENIX resulting in a final R_{factor} of 17.6% and R_{free} of 26.3%. Water molecules were added using unrestrained and restrained refinement and the average B-factor of the entire protein was 47 (\AA^2). The root mean square deviations (RMSD) from ideal geometry are 0.013 \AA for bond lengths and 1.24 \AA for bond angles. Crystal structure data and refinement statistics of CVCP are reported in Table 3.1.

Table 3.1: Data collection and refinement statistics for CVCP

<u>Crystallographic Data</u>	
Space group	<i>P1</i>
Cell dimensions (Å)	<i>a</i> =33.3, <i>b</i> =39.1, <i>c</i> =59.3
Resolution range (Å)	50-2.2
Completeness (%) (Last Shell)	91.1 (74.9)
<i>R</i>_{sym} (%)^a (Last Shell)	0.06 (0.22)
Mean <i>I</i>/<i>σ</i>(<i>I</i>) (Last Shell)	14.4 (4.2)
No. of unique reflections	13451
Molecules per asymmetric unit	2
Matthews coefficient (Å³ Da⁻¹)	2.29
Solvent content (%)	49.0
Multiplicity (Last Shell)	2.4 (2.2)
Wilson B-factor	32.9
<u>Refinement</u>	
No. of Residues	302
Water molecule	88
<i>R</i>_{cryst} (%)	17.6
<i>R</i>_{free} (%)	26.3
Average <i>B</i>-factor (Å²)	47.0
r.m.s.d on bond lengths (Å)	0.013
r.m.s.d on bond angles (Å)	1.24
<u>Ramachandran plot (%)</u>	
Favoured	97
Allowed	3
Outliers	0

$$^a R_{sym} = \frac{\sum_{hkl} \sum_{i=1}^n |I_{hkl,i} - \overline{I_{hkl}}|}{\sum_{hkl} \sum_{i=1}^n I_{hkl,i}}$$

3.4.2 Monomer architecture

Similar to other alphavirus CPs structures, the overall assembly of protease domain of CHIKV consists of two β -barrel sub-domains. The sub-domain 1 comprises of five β -strands and sub-domain 2 comprises of seven β -strands. Both the subdomains are linked through a 10 residues long linker (Ala175-Glu184) (Fig. 3.4.2.1). The β -sheets of N-terminal sub-domain 1 [β 1(114–119), β 2 (122–130), β 3 (133–137), β 4 (155–157), β 5 (161–166)] and C-terminal sub-domain 2 [β 6 (185–188), β 7 (193–197), β 8 (200–204), β 9 (214–219), β 10 (222–233), β 11 (237–244), β 12 (251–253)] arrange to form a Greek key motif, a key feature of chymotrypsin-like serine proteases. In addition to this, the N-terminal sub-domain 1 also contains two helices: α 1 (147-150) and α 2 (169-174) (Fig. 3.4.2.1). Each sub-domain is stabilized by the presence of salt bridges. Five salt-bridges (Glu116-Lys118, Lys118-Asp145, Lys136-Glu163, His139-Asp161, and Asp148-Lys151) are confined to sub-domain 1 whereas four salt bridges (His180-Asp248, Arg200-Glu234, Arg215-Asp212 and Arg223-Asp219) are confined to sub-domain 2 (Table 3.2). These salt bridges are responsible for stability and compact sub-domain structure of CVCP. The C α RMS deviation between both CVCP monomers A and B was 0.24 Å.

Table 3.2. CVCP intra-subdomain salt bridges

Subdomain I (residues 114-174)		
Residue 1	Residue 2	Distance (Å)
OE1; Glu116	NZ; Lys118	3.9
NZ; Lys118	OD2; Asp145	3.6
NZ; Lys136	OE1; Glu163	2.5
ND1; His139	OD1; Asp161	3.6
OD1; Asp148	NZ; Lys151	3.8
Subdomain II (residues 185-261)		
Residue 1	Residue 2	Distance (Å)
ND1; His180	OD2; Asp248	2.4
NH1; Arg200	OE1; Glu234	3.4
NH2; Arg215	OD1; Asp212	3.8
NH2; Arg223	OD2; Asp219	3.0

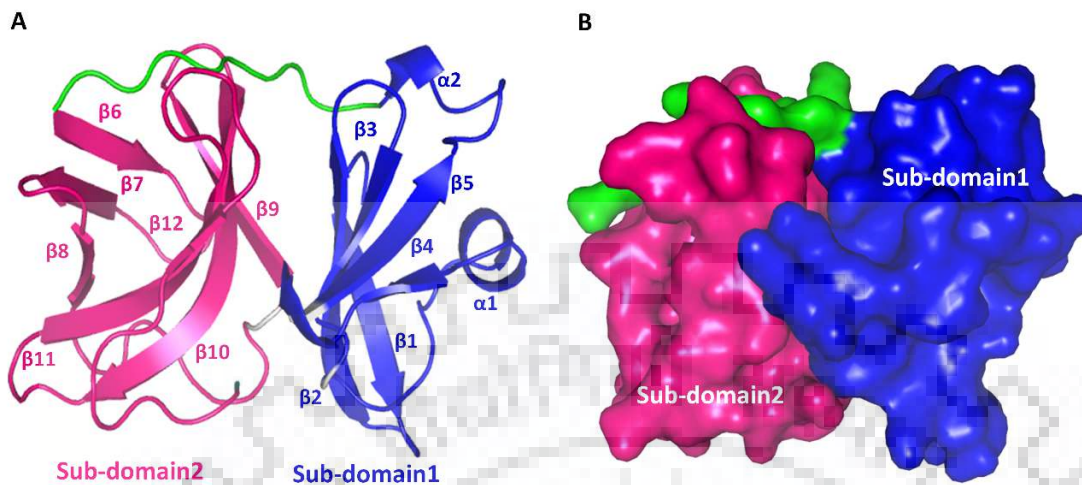


Figure 3.4.2.1: Crystal structure of CVCP (A) Cartoon representation of the CVCP crystal structure with detail secondary structure elements of two sub-domains. The sub-domain 1, sub-domain 2 and the linker region are represented in blue, magenta and green color, respectively. (B) Surface representation of the sub-domain 1, sub-domain 2 and the linker region of CVCP crystal structure. The color scheme is same as in the cartoon view representation.

3.4.3 Substrate binding site architecture

The spatial architecture of the substrate binding pocket is composed of conserved catalytic triad of serine protease family including alphavirus CPs. The conserved catalytic triad of CVCP consists of His139, Asp161 and Ser213 residues and is positioned in between the two sub-domains (Fig. (Fig. 3.4.3.1). Additionally, the CVCP is also composed of ${}_{211}\text{GDSG}_{214}$ motif having key active site serine residue (Fig. 3.4.4.3). This feature is conserved among alphavirus CPs and other members of chymotrypsin-like serine proteases. The residues Ser213 and Gly216 are involved in forming the oxyanion hole. Alphavirus capsid protease cleaves the scissile bond between the W/S site and leaves the terminal Trp261 residue in the substrate binding pocket further inactivating the capsid protease.

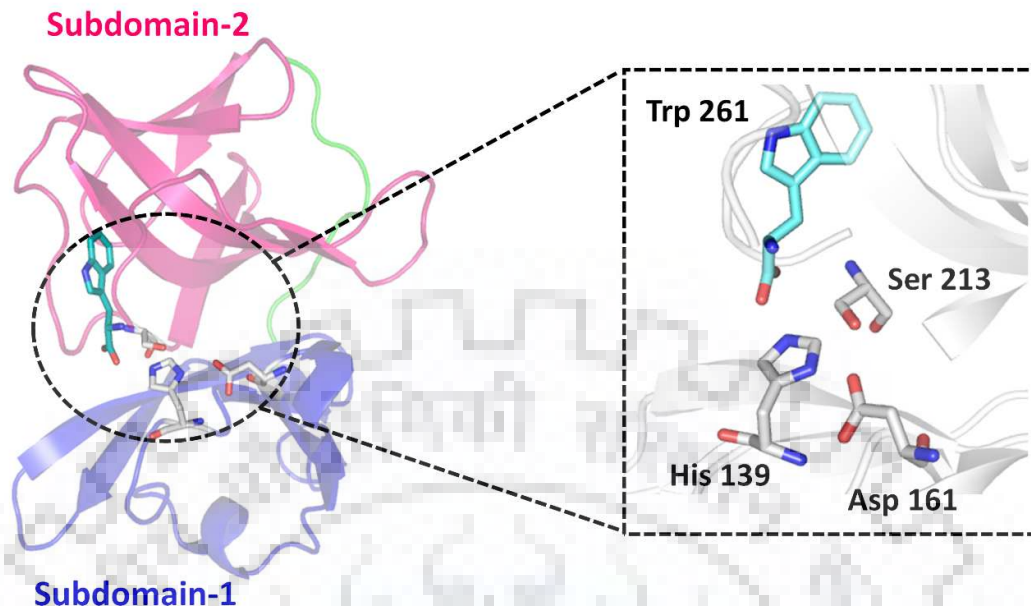


Figure 3.4.3.1: Catalytic architecture of CVCP. The conserved catalytic triad of CVCP protease is present in the cleft between the two sub-domains. The conserved catalytic triad residues (His139, Asp161 and Ser213) are shown in sticks (zoom in view). The terminal tryptophan residue (Trp261) left after the auto *cis*-proteolytic cleavage remains bound to the active site and blocks further protease activity.

3.4.4 Dimer architecture

CVCP has monomeric nature in solution and crystallographic dimer per asymmetric unit in crystal form. Previous studies have reported the crystallographic dimer in the crystal structures of Sindbis CP (SCP) (PDB: 1WYK), Aura CP (AVCPΔ2) (PDB: 4UON) and Semliki Forest CP (SFCP) (PDB: 1VCP) (Aggarwal et al., 2014; Choi et al., 1997; Choi et al., 1991). The crystallographic dimeric interface in CVCP crystal structure is found to be similar as observed in SFCP, AVCPΔ2 and SCP crystal structures. The crystallographic dimer in CVCP has contacts through the C-terminal subdomains (tail-to tail) from each monomer (Fig. 3.4.4.1). However, in SFCP a head-to-head dimeric contact, involving the interaction between the N-terminal subdomains exists as BC dimer.

The AB dimeric interface area in CVCP was found to be $\sim 441.6 \text{ \AA}^2$, encompassing 5.5% of total solvent-accessible area (SAA) as calculated using PISA. The previous reports show that the dimeric interface area in SCP, AVCPΔ2 and SFCP are $\sim 409 \text{ \AA}^2$, $\sim 375 \text{ \AA}^2$ and $\sim 374 \text{ \AA}^2$, respectively. This suggests that in CVCP, a large part of C-terminal sub-domains from each monomer is involved in dimeric interface interactions. Both the monomers contribute two loops (183PE₁₈₄ and 220NK₂₂₁) and two β -sheets (185GYYN₁₈₈ and 193AVQYS₁₉₇) for monomer-monomer interactions. The contacts between two monomers are facilitated by involvement of 13

residues from each monomer. The dimer interface include residues Pro183, Glu184, Gly185, Tyr186, Tyr187, Asn188, Trp189, Gly192, Ala193, Asp219, Asn220, Lys221 and Arg223 from each monomer. The dimer stability can be determined by interactions taking place at the interface. The residues Tyr186, Asn188 and Asn220 of monomer A are involved in forming H-bonds with Tyr186 and Asn220 of monomer B (Fig. 3.4.4.2 and Table 3.3). Apart from this, hydrophobic interactions also play a crucial role in CVCP dimer stability.

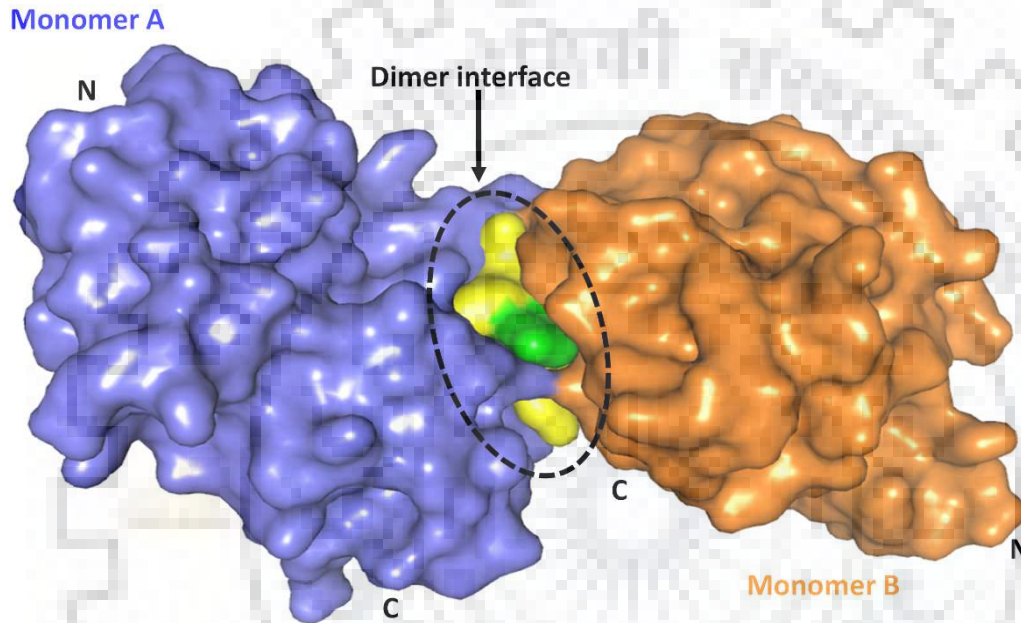


Figure 3.4.4.1: Dimer interface interactions: Surface view of monomer A and B represented in blue and orange color respectively. The residues involved in hydrogen bond interactions at the dimer interface are represented in yellow (monomerA) and green color (monomerB), respectively

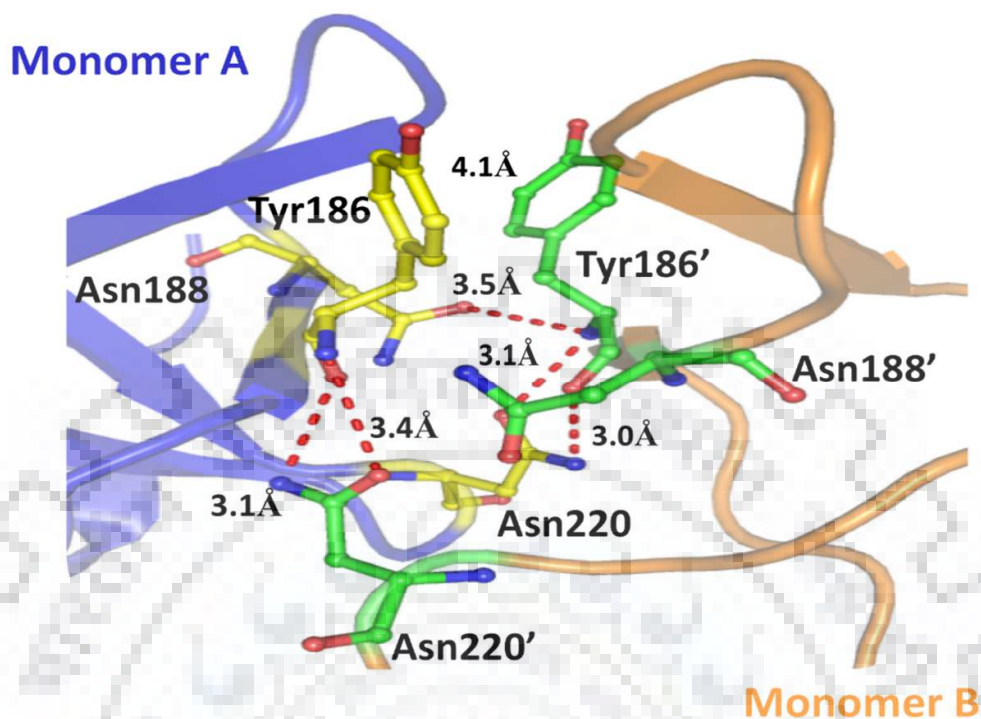


Figure 3.4.4.2: Structural insights of dimer interface interactions: Zoom in view of CVCP dimeric interaction. Monomer A and B are shown in blue and orange color, respectively. The residue Tyr186 of one monomer forms hydrogen bonds with Asn188 and Asn220 of other monomer. The residues involved in polar interactions of monomer A and B are represented in yellow and green color sticks, respectively. Monomer B residues are represented with a prime (') sign.

Table 3.3. Interface polar interactions between the two crystallographic monomers of CVCP

Monomer A	Monomer B	Distance (Å)
O; Tyr186	ND2; Asn220	3.1
N; Tyr186	OD1; Asn220	3.4
OD1; Asn188	N; Tyr186	3.5
OD1; Asn220	N; Tyr186	3.1
ND2; Asn220	O; Tyr186	3.0

The region from 183Pro-Glu-Gly-Tyr-Tyr-Asn188, and residues Gly192, Ala193 and Asn220) are highly conserved among CVCP, SFCP, SCP and AVCPΔ2 (Fig. 3.4.4.3). However, the Tyr186 in CVCP is replaced by His192, Phe188 and Phe191 in SFCP, SCP and AVCPΔ2, respectively (Fig. 3.4.4.3). Additional mutational and structural studies need to be carried out for analysing the biological significance of the conserved residues at the dimer interface.

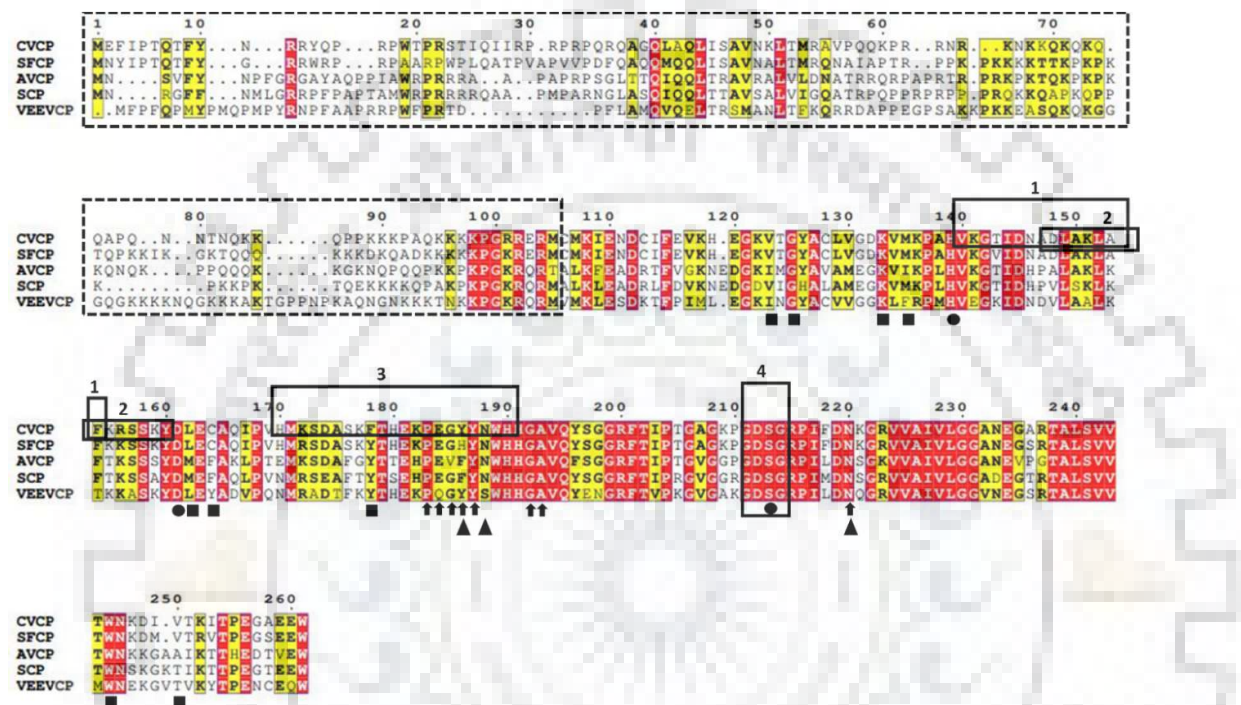


Figure 3.4.4.3: Multiple sequence alignment of CVCP with CP's from other alphaviruses:

The N-terminal domain of CHIKV capsid protein is shown with dotted box. The conserved residues are highlighted in red background. The solid black color squares under the amino acids represented the residues of the hydrophobic pocket involved in interaction with dioxane, PCA, MDA and EAB. The highly conserved catalytic residues (His139, Asp161 and Ser213 in CVCP) of CP's are represented by solid black circles below the residues. The conserved amino acid residues involved in crystallographic dimer formation are represented by upward black arrows below the residues. The residues involved in polar interactions at the dimeric interface are represented by solid black triangles below the residues. The predicted NES and epitopic regions I and II and GDSG motif (211GDSG214) are shown in the box 1, 2, 3 and 4, respectively.

3.4.5 CVCP-cdE2 glycoprotein interactions

The cryo-EM structure of VEEV at a resolution of 4.4 Å (Zhang et.al), provided the structural insights of viral surface assembly for mature virion formation (Zhang et al., 2011). The VEEV structural analysis revealed the conserved hydrophobic pocket on CP surface as the site for molecular contacts with cdE2. The CHIKV E2 is 423 amino acids long protein and consists of three regions, an ectodomain of 364 residues followed by transmembrane helix (TM) of 26 residues and a 33 residue cdE2 domain. Additionally, cdE2 is associated with two N-linked

glycans at positions Asn263 and Asn345 (Sun et al., 2013). Though, the exact CP-cdE2 interaction is not clear until now. To study the CP-cdE2 interactions taking place at the CHIKV viral surface, the CVCP crystal structure and the modeled cdE1 and cdE2 were fitted into VEEV cryo EM density map (Fig. 3.4.5.1).

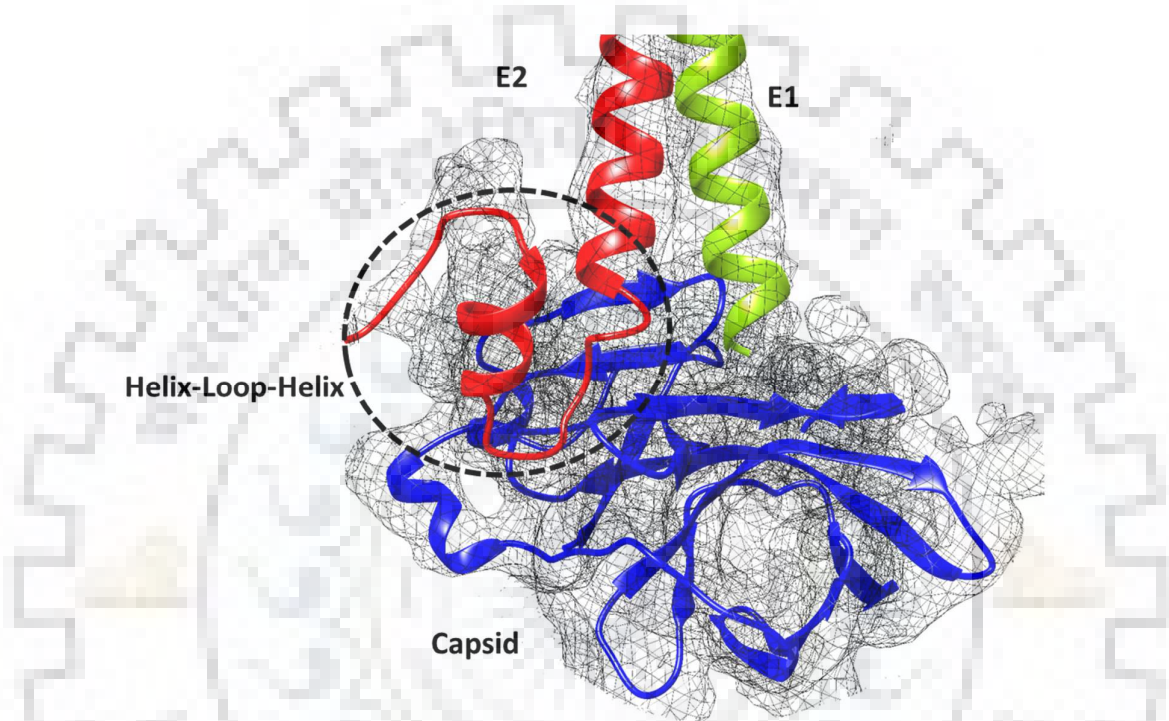


Figure 3.4.5.1: CVCP crystal structure and E1, E2 glycoprotein homology models fitted in to cryo-EM density map of VEEV (EMDB ID: 5275): (A) Depiction of CVCP crystal structure and E1, E2 homology models fitted in to cryo-EM density map (grey color) of VEEV. The CVCP, E1 and E2 are represented by blue, green and red color, respectively. (B) The helix-loop-helix motif (enclosed in dotted circle) of cdE2 fits in to CVCP hydrophobic pocket is also shown.

The cdE2 is highly conserved among the alphavirus family and through its H-L-H conformation, it fits into CVCP hydrophobic pocket and interacts with the surrounding residues of the pocket (Fig. 3.4.5.2).

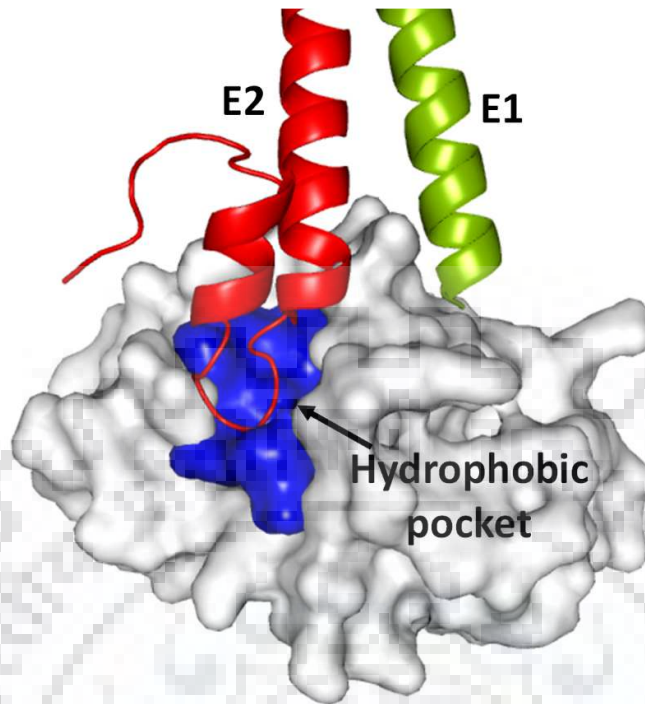


Figure 3.4.5.2: Depiction of Helix-loop-Helix of cdE2 glycoprotein fitted in to CVCP hydrophobic pocket: Surface view representation of CVCP highlighting hydrophobic pocket (blue color) and fitted loop of helix-loop-helix (cartoon) of cdE2 is shown.

A conserved Tyr residue (Tyr400 in CHIKV) is located at the end of the helix-I of the H-L-H motif of cdE2. The studies in Sindbis virus proposed that this conserved Tyr400 residue of cdE2 is responsible for making contacts with CP (Lee et al., 1996; Sun et al., 2013). Conversely, in CHIKV cdE2 the Tyr400 is aligned away from the CP hydrophobic pocket and exhibit the hydrophobic contacts with Leu411 and Leu412 of another helix (helix-II) of H-L-H motif of cdE2 (Fig. 3.4.5.3).

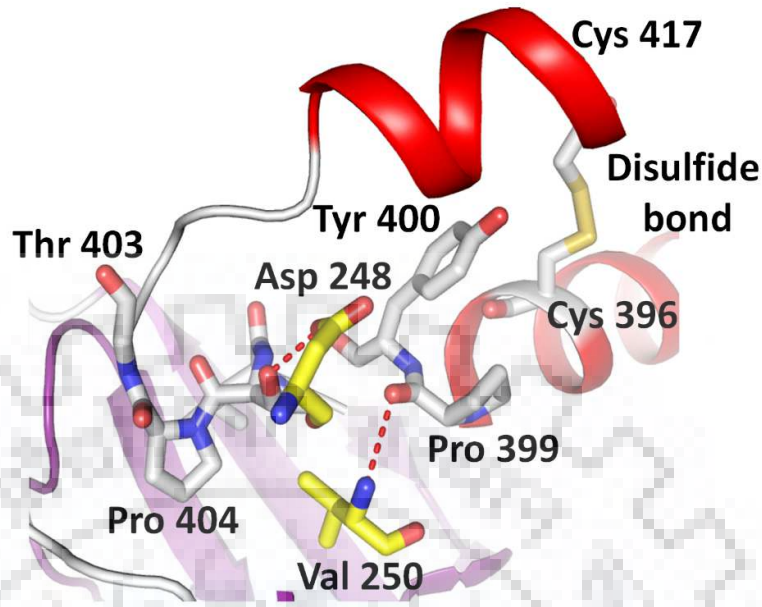


Figure 3.4.5.3: CVCP- cdE2 interactions: The critical residues involved in the capsid-cdE2 interactions are shown in yellow and grey color sticks, respectively. The polar interactions are represented in red dotted lines and the cysteine disulfide bond with in helix-loop-helix motif of cdE2 is shown in yellow color. The proline residue (Pro 404) of cdE2 bound to the CVCP hydrophobic pocket is also shown.

Structural analysis of CHIKV cdE2 unveils three conserved cysteine residues (Cys396, Cys416, and Cys417) in the helix-loop-helix motif (Fig. 3.4.5.4). The Cys396 residue of helix-I of helix-loop helix motif forms disulfide bond with Cys417 lying at the loop region at the end of helix-II (Fig. 3.4.5.3).

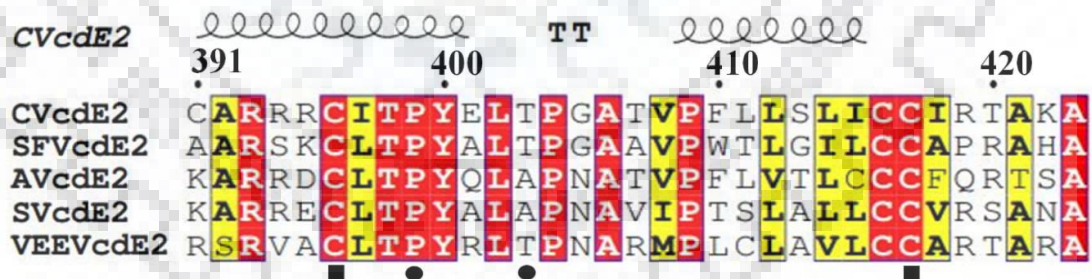


Figure 3.4.5.4 Multiple sequence alignment of cytoplasmic domain of E2 glycoprotein (cdE2) from different alphaviruses: {Chikungunya virus (CVcdE2), Semliki Forest virus (SFVcdE2), Aura virus (AVcdE2), Sindbis virus (SVcdE2) and Venezuelan Equine Encephalitis virus (VEEVcdE2)}. The cdE2 domain (residues 391-423) of the CHIKV has helix-loop-helix motif as represented in the multiple sequence alignment. The conserved residues are highlighted in red background. The conserved residues of cdE2 involved in the formation of disulfide bond are shown with square below the residues. The cdE2 residues involved in the polar interactions with CVCP are shown with solid circles under the residues.

These two residues have been suggested to undergo palmitoylation (typical feature of membrane proteins) that helps in anchoring and orienting of cdE2 to the inner side of membrane that facilitate the interactions with CP (Jose et al., 2012). Similar orientations of Tyr400 residue and disulfide bond (Cys396-Cys417) are reported in VEEV and Aura virus cdE2 (Aggarwal et al., 2012; Zhang et al., 2011). Mutations of these conserved Cys residues results in virus assembly defects and affects the entry of virus in host cells, thereby, decreasing the virus infection (Jose et al., 2012). The loop region of the H-L-H motif of cdE2 contains seven residues (402Leu-Val408). The residues of the loop region of cdE2 interact with residues of the hydrophobic pocket of CVCP. The Pro404 lying in the loop region gets deep buried into the hydrophobic pocket and makes hydrophobic contacts with the surrounding residues (Fig. 3.4.5.5). In the crystal structure of CVCP, ethylene glycol was found in the hydrophobic pocket, inhabiting the identical location as Pro404 of cdE2.

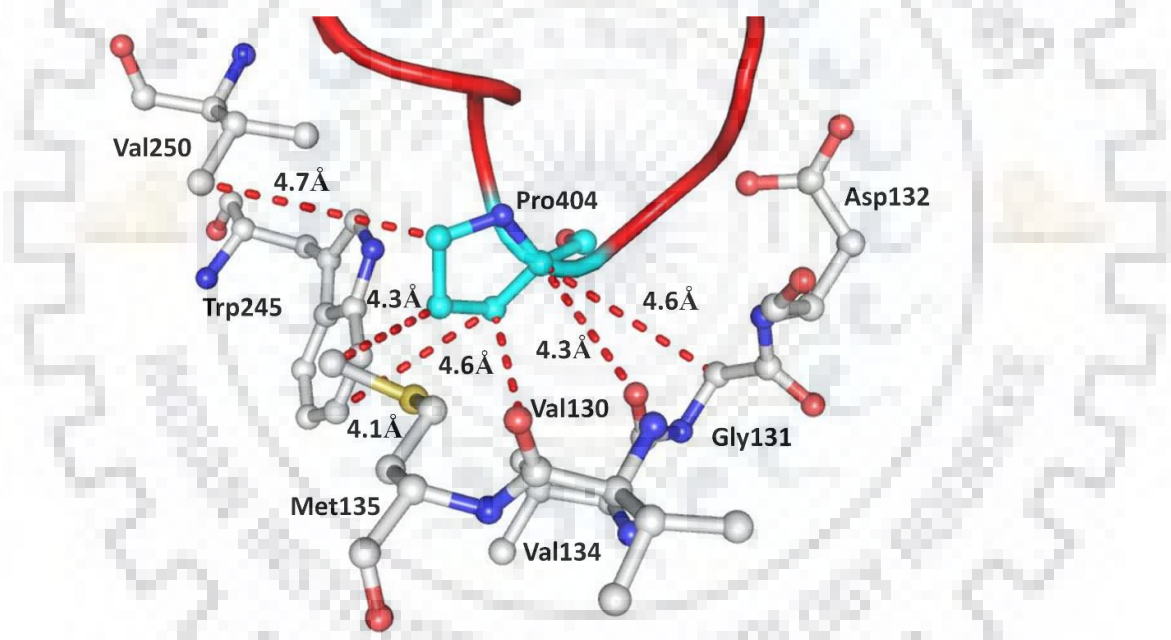


Figure 3.4.5.5: Hydrophobic interactions of Pro404 of cdE2 with CVCP hydrophobic pocket: The residues of CVCP involved in the hydrophobic interactions with Pro404 of cdE2 are shown in grey color sticks. The hydrophobic interactions are represented in red dotted lines. The proline residue (Pro404) of cdE2 bound to the CVCP hydrophobic pocket is shown in cyan color.

3.4.6 Structural comparison of CVCP with other alphavirus CP's

Structural comparison of CVCP with the known crystal structures of alphavirus capsid proteases was done for detail structural analysis. The sequence based homology search for CVCP against PDB database using NCBI BLASTp tool mostly gave hits of alphaviral capsid crystal structures (ranges from 68 to 93% identity). Structural homolog search using DALI server

revealed the highest structural similarity of CVCP with SFCP (PDB: 1VCP) having identity 93 %, Z-score 29.2 and RMSD 0.7 Å, followed by AVCP (PDB: 4AGJ, with identity 69 %, Z-score 28.7 and RMSD 0.9 Å), SCP (PDB: 1WYK, with identity 67 %, Z-score 28.5 and RMSD 0.9 Å) and VEEVCP (PDB:1EP5, with identity 63 %, Z-score 27.8 and RMSD 0.9 Å). Following the highest structural similarity with Alphavirus CP's (SFCP, AVCP, SCP and VEEV), DALI server also revealed structural similarity to the other members of serine protease family like chymotrypsin, flavivirus NS3 protease, 3C like proteinase, glutamyl endopeptidase, chymase and elastase. Structural similarity of CVCP with flavivirus protease indicates the evolution of flavivirus and alphavirus protease from a common ancestry of chymotrypsin-like serine proteases. Superimposition of CVCP with related alphavirus CP structures and structural homology results from DALI reveals the β -barrel structure is highly conserved among various members of serine protease family. Although the overall structural architecture and active site conformation of alphavirus enzymes is typically conserved, slight variation in the position of three loop regions lying on the surface of the protein was observed. Figure 3.4.6.1 depicts the three loop regions showing high C α RMS deviation. Region I lies in-between the β 1 and β 2 and involves residues His119-Lys122 (Circle 1). This region in SFCP has been reported to be involved in making head-to-head contacts in SFCP. Region II is part of the linker-loop connecting the β -sheets of two sub-domains. This region is of variable length and embraces diverse conformations. The length of this inter-domain linker-loop in CVCP, AVCP, SFCP, SCP and VEEVCP is of 10, 18, 10, 13 and 11 amino acid residues long stretch, respectively (Circle 2). A part of this inter-domain loop region contributes to formation of the conserved hydrophobic pocket where the tail of E2 glycoprotein binds. The region III is the loop region connecting β 11 (237–244) and β 12 (251–253). The length of this loop region is six residues long in CVCP (residues: 245-250), whereas in SFCP the length of this loop is 2 residues (residues: 253-254). When compared to the other alphavirus CP's, this loop region was longer in AVCP, VEEVCP and comparable to SCP (Circle 3). This flexible loop region showing maximum conformation variation in CVCP interacts with cdE2 glycoprotein and is also involved in dimer interface contacts (Choi et al., 1997).

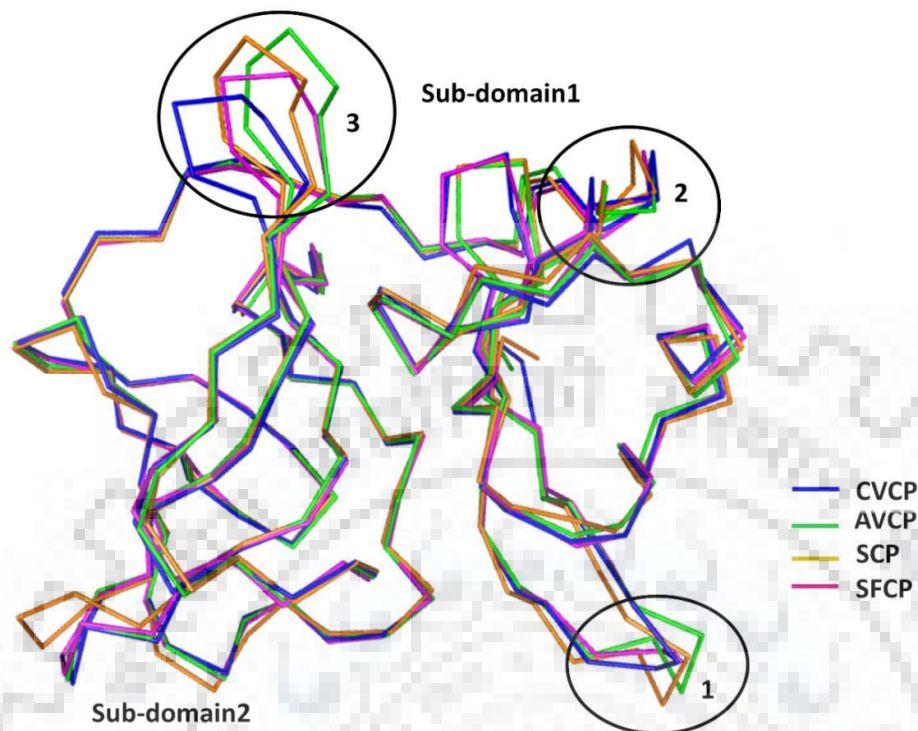
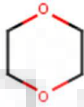
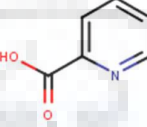
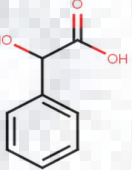
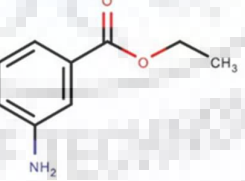


Figure 3.4.6.1: Superimposition of CP's from different alphaviruses: Structural alignment of CVCP (blue) with AVCP (green), SCP (orange) and SFCP (magenta). The three loop regions showing high structural variation among CPs are enclosed with in three circles (Circle 1-3).

3.4.7 Identification and molecular docking of small heterocyclic molecules

The crystal structural studies of capsid protein in complex with dioxane (SCP, PDB ID: 1WYK) and AVCP PDB ID: 4AGJ) (Aggarwal et al., 2012; Choi et al., 1996) and studies of alphavirus inhibition by heterocyclic compounds targeting the conserved hydrophobic pocket (Aggarwal et al., 2017; Aggarwal et al., 2012; Sharma et al., 2016) have revealed that the hydrophobic pocket of CP is a potential antiviral drug target. Therefore, in this study PubChem database was searched for drug-like heterocyclic compounds similar to dioxane, picolinic and piperazine. The small molecules containing heterocyclic ring were identified for potentially targeting the CP-cdE2 interactions. These small molecules were docked into the hydrophobic pocket of CVCP crystal structure. Based on the docking score, MDA and EAB showed better binding tendency to CVCP hydrophobic pocket than dioxane and PCA. Table 3.4 shows the molecular structure and docking score of the compounds.

Table 3.4: The molecular docking results of dioxane, picolinic acid, (S)-(+)-Mandelic acid and Ethyl 3-aminobenzoate binding to CVCP hydrophobic pocket

Ligand/ Molecular formula	Molecular Structure	Docking Score (kcal/mol)
Dioxane C ₄ H ₈ O ₂		-2.88
Picolinic Acid (PCA) C ₆ H ₅ NO ₂		-3.59
(S)-(+)-Mandelic acid (MDA) C ₈ H ₈ O ₃		-4.03
Ethyl 3-aminobenzoate (EAB) C ₇ H ₇ NO ₂		-3.93

Detail molecular interaction of MDA and EAB compounds in the hydrophobic pocket of CVCP is shown in Figure 3.4.7.1 and 3.4.7.2. The binding affinities of these two molecules were further experimentally evaluated using SPR and fluorescence spectroscopy.

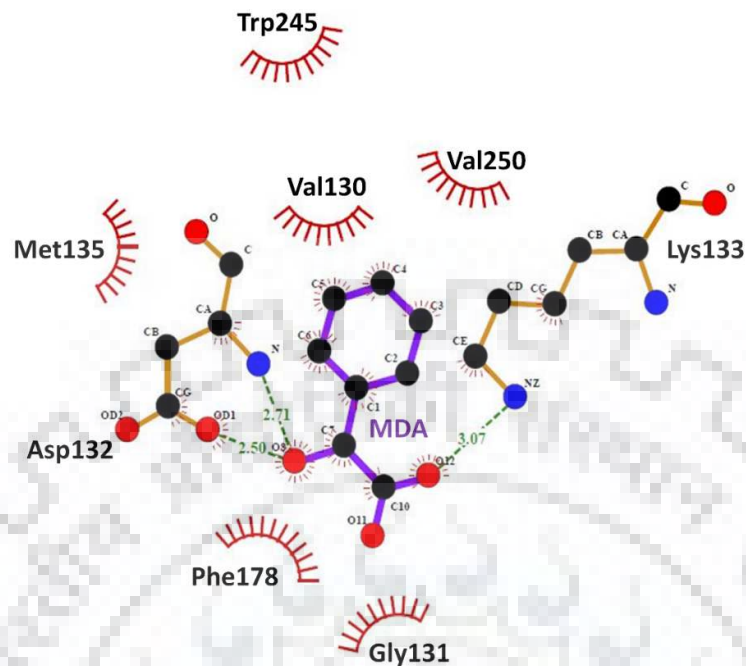


Figure 3.4.7.1: Binding analysis of docked (S)-(+)-Mandelic acid (MDA) into CVCP hydrophobic pocket: The diagram showing the binding of MDA in the hydrophobic pocket of CVCP. The residues of the hydrophobic pocket of CVCP involved in hydrophobic interactions with MDA are shown in red color, while the MDA is shown purple color. The polar interactions of MDA with the hydrophobic pocket residues are shown in green dashes.

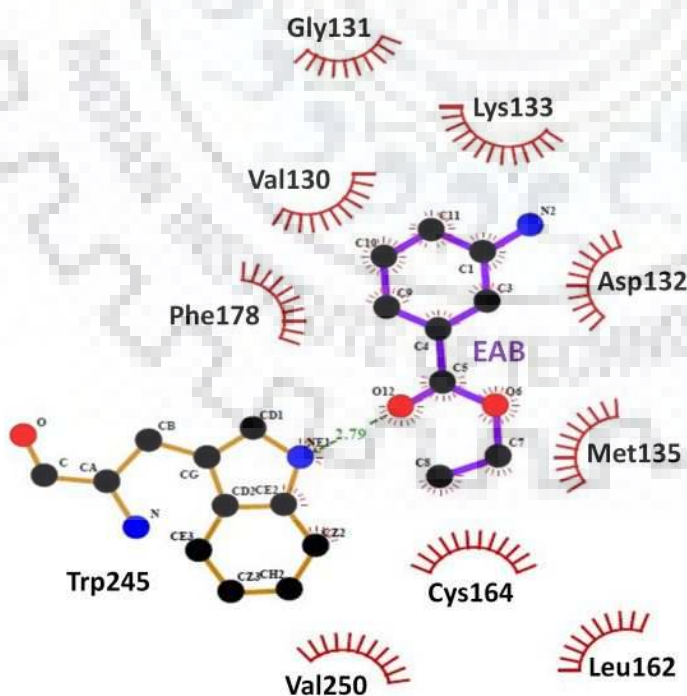


Figure 3.4.7.2: Binding analysis of docked Ethyl 3-aminobenzoate (EAB) into CVCP hydrophobic pocket: The diagram showing the binding of EAB in the hydrophobic pocket of CVCP. The residues of the hydrophobic pocket of CVCP involved in hydrophobic interactions with EAB are shown in red color, while the EAB is shown purple color. The polar interactions of MDA with the hydrophobic pocket residues are shown in green dashes. MDA and EAB binds to the CVCP hydrophobic pocket more strongly than dioxane and picolinic acid.

3.4.8 Surface plasmon resonance

The ligand (CVCP protein) was immobilized on the flow cell 4 (FC-4) and the final immobilization level was 2,771 RU. Flow cell 3 (FC-3) was selected as a reference to minimize the unwanted drift and systematic noise. For binding studies, the identified MDA and EAB molecules were injected over the immobilized CVCP, at concentrations ranging from 0.0625 mM to 1 mM and 0.125 mM to 1mM, respectively. The responses achieved upon binding of the two analyte, from different concentration range of MDA and EAB were used to measure the binding kinetic parameters for the interaction with CVCP. For kinetic studies, the contact time of MDA and EAB binding to the CVCP protein was monitored for 180 s at 30 μ l/min flow rate. The responses of reference cell obtained for the control buffer blanks were flat, demonstrating the stability of the base line during the experiment.

Kinetic parameters for binding of MDA and EAB to the immobilized CVCP protein were measured. The data obtained was evaluated by the Biacore Evaluation software (Biacore T200 Evaluation Software, Version: 2.0) (Fig. 3.4.8.1 and 3.4.8.2). The data was fitted as per the global fitting of the kinetic model with two state model. The values of the dissociation constants (K_D) obtained from SPR for binding of MDA and EAB to CVCP were 1.2×10^{-3} M and 0.2×10^{-9} M respectively.

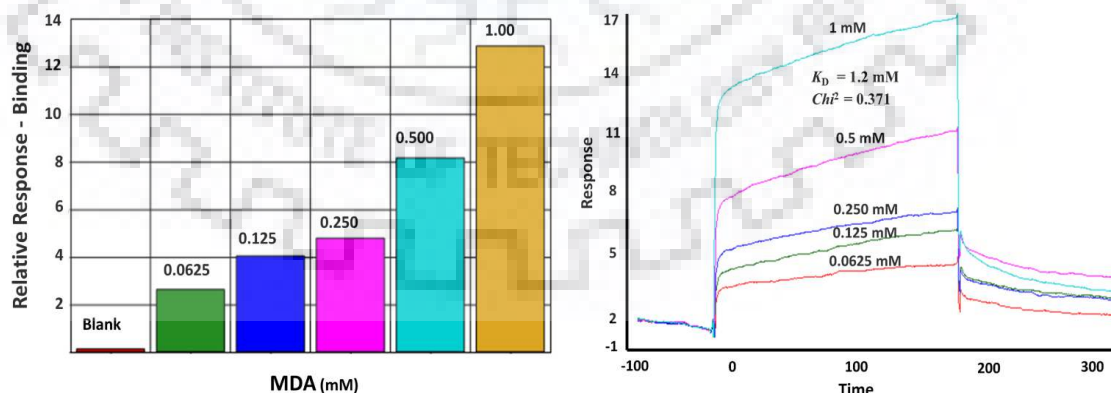


Figure 3.4.8.1: SPR binding and kinetic analysis of (S)-(+)-Mandelic acid (MDA): SPR binding and kinetic analysis of (S)-(+)-Mandelic acid (MDA) have been performed. The relative responses obtained for the binding of MDA to the immobilized CVCP protein are shown. The data was fitted by two state binding model with Chi^2 value of 0.37. The K_D value obtained for the binding of MDA to immobilized CVCP was 1.2×10^{-3} M.

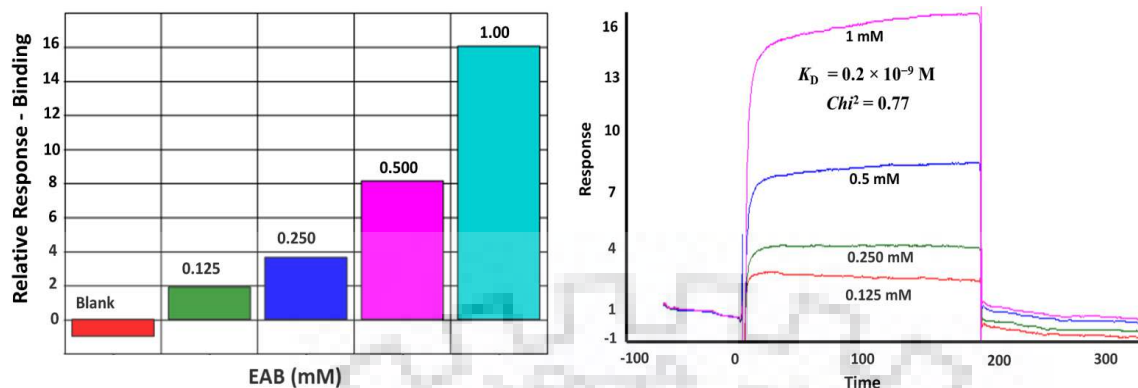


Figure 3.4.8.2: SPR binding and kinetic analysis of Ethyl 3-aminobenzoate (EAB): The relative responses obtained for the binding of Ethyl 3-aminobenzoate (EAB) to the immobilized CVCP protein are shown. The data was fitted by two state binding model with Chi² value of 0.77. The KD value obtained for the binding of EAB to immobilized CVCP was 0.2×10^{-9} M.

3.4.9 Effect of MDA and EAB on CVCP intrinsic fluorescence

The intrinsic fluorescence of CVCP was measured in its native form (residues 106-261) and upon binding of MDA and EAB at varied concentrations. As suggested from docking studies of dioxane, MDA and EAB are predicted to bind the hydrophobic pocket of CVCP with higher affinity than dioxane and PCA. The binding of MDA and EAB to CVCP hydrophobic pocket results in decrease of fluorescence intensity with the increase in concentration as shown in Fig. 3.4.9.1 and Fig. 3.4.9.2. Both the MDA and EAB quenches the CVCP fluorescence and shift the emission maxima to the longer wavelengths i.e exhibiting red shift. It is suggested that the binding of MDA and EAB in the CVCP hydrophobic pocket further exposed the buried Trp to the solvent and results in red shift in the emission spectra. Fluorescence studies further validated the SPR results that MDA and EAB bind to the hydrophobic pocket of CVCP.

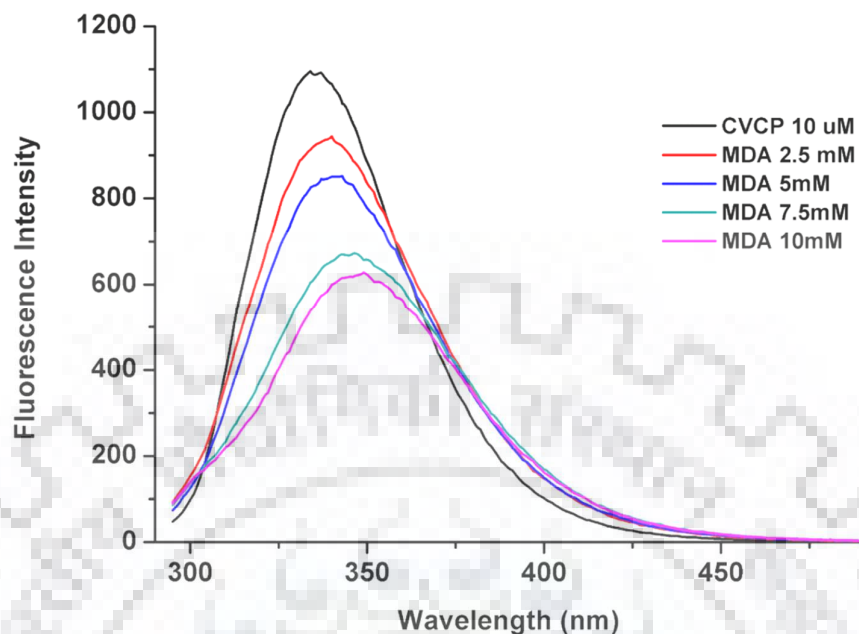


Figure 3.4.9.1: Intrinsic fluorescence spectroscopy analysis of MDA: The intrinsic fluorescence of CVCP native protein and upon binding to MDA was determined. The emission spectra was recorded upon addition of MDA at different concentrations up to 10 mM (2.5, 5.0, 7.5 and 10 mM) to the native CVCP (10 μ M), after incubation for 15 min. The fluorescence intensity got decreased with the increase in concentration of MDA.

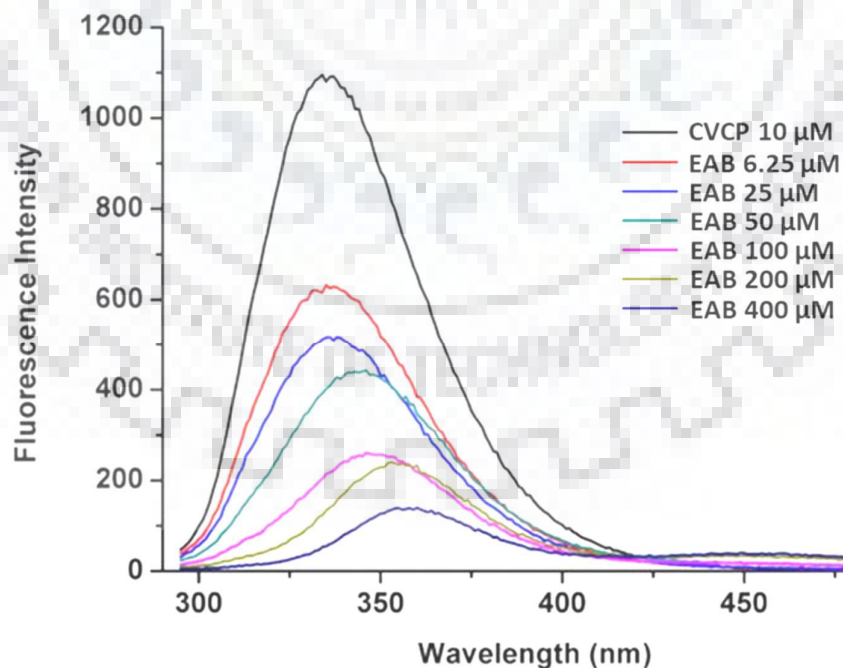


Figure 3.4.9.2: Intrinsic fluorescence spectroscopy analysis of EAB: The intrinsic fluorescence of CVCP native protein and upon binding to EAB was determined. The emission spectra was recorded upon

addition of EAB at different concentrations up to 400 μM (6.25, 25, 50, 100, 200 and 400 μM) to the native CVCP (10 μM), after incubation for 15 min. The fluorescence intensity got decreased with the increase in concentration of EAB.

3.4.10 Epitope identification

The crystal structure of CVCP was used for identifying and predicting linear B-cell epitopes using in silico techniques like Kolaskar and Tongonkar antigenicity prediction tool, Bepipred and ABCpred (Fig. 7). The regions $_{148}\text{DLAKLAFKRSSKY}_{160}$ and $_{170}\text{HMKSDASKFTHEKPEGYYNWH}_{190}$ were identified as the probable epitopic regions involved in causing antigenic reaction in the host (Fig. 3.4.4.3 and Fig. 3.4.10.1). The hydrophilic and solvent accessibility of predicted mobile regions are important parameters contributing towards antigenicity of a peptidic fragment. Therefore, the identified peptide epitopes were further evaluated for their biochemical properties like surface accessibilities, flexibility, hydrophobicity and hydrophilicity. These regions lie at the surface of the CVCP, the region I (148-160) is a long stretch containing the loop regions and include the $\beta 4$ and the helix region; while region II (170-190) contain flexible linker-loop region and $\beta 6$. The region I also include a part of the NES region of the capsid protein which is recognized by the exportin (CRM1) of the host. The identified epitopes of CHIKV CP (148-160 and 170-190) can potentially be utilized for developing epitope based vaccines against chikungunya disease.

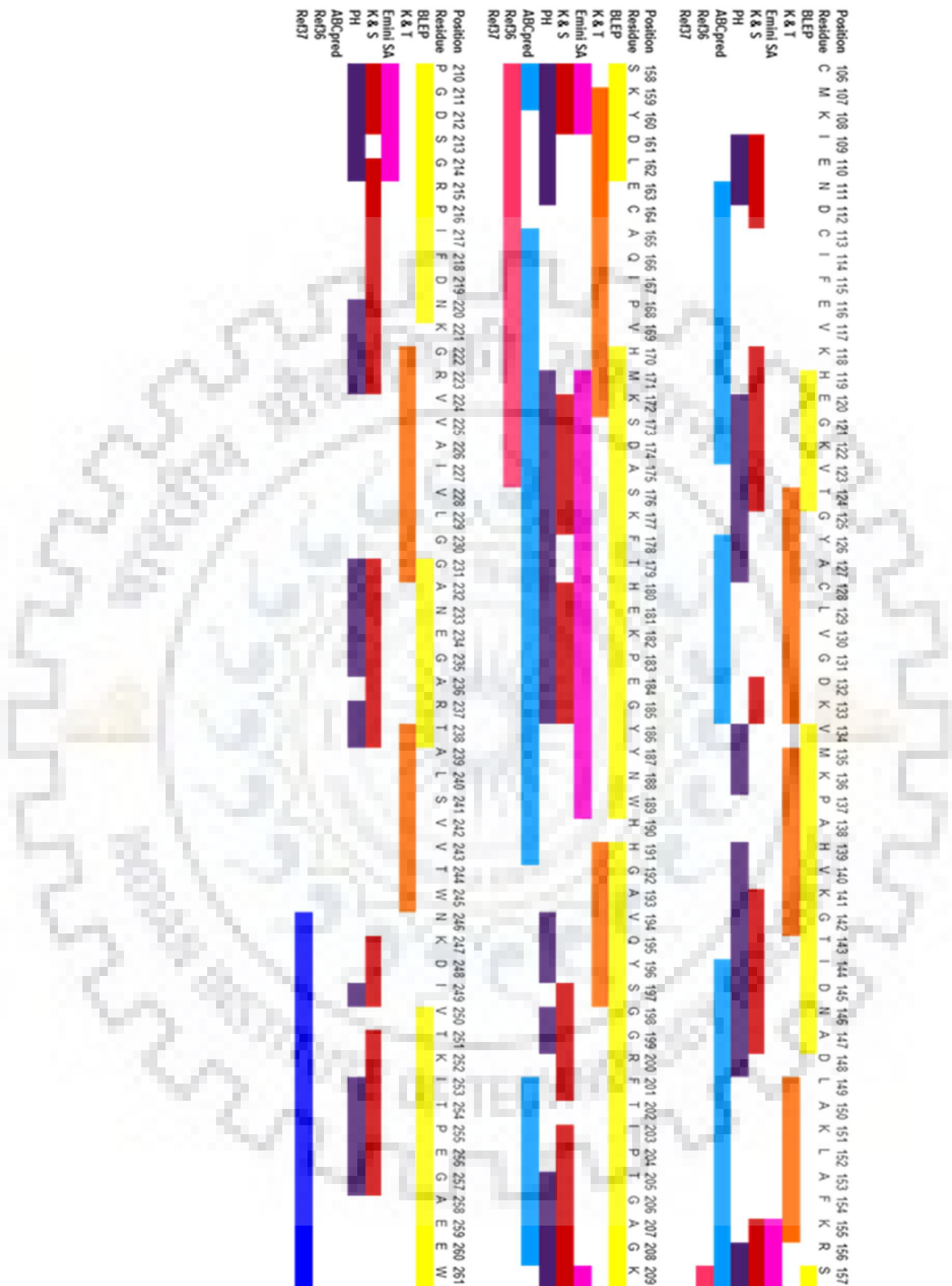


Figure 3.4.10.1: In-silico epitope identification: Location of the identified epitopes on the protease domain of CHIKV capsid protein predicted using in silico tools have been depicted by various colors. Yellow- Bepipred Linear Epitope Prediction (BLEP), orange- Kolaskar and Tongonkar antigenicity prediction tool (K&T), magenta- Emini Surface Accessibility Prediction (Emini SA), red- Karplaus & Schulz Flexibility Prediction (K & S Score), violet- Parker Hydrophobicity Prediction (PH), cyan-

ABCpred. The epitopes already reported in literature (Ref. 36, 37) has been represented in pink and blue color, respectively.

3.5 Conclusion

Alphavirus capsid protein is a multifunctional protein, divided into two domains: RNA binding N-terminal domain and autoproteolytic C-terminal domain. The present study, describes the crystal structure of CVCP at 2.2 Å resolution and showed that the protease domain has two β-barrel sub-domains, here referred to as N-terminal sub-domain 1 (residue 111-174) and C-terminal sub-domain 2 (residue 185-261) connected together through a 10 residue long linker (residue 175-184). Belonging to the serine proteases, CVCP has a conserved catalytic triad which contains His139 and Asp161 (contributed by sub-domain 1) and Ser213 (contributed by sub-domain 2) residues. The catalytic triad is positioned in between two sub-domains. The key catalytic Ser213 residue along with Gly211 helps in stabilization of the backbone oxygen atom of Trp261 thereby leading to formation of oxyanion hole. The *cis*-proteolytic activity of the CP is responsible for cleavage from the nascent structural polyprotein during translation (peptide bond cleavage between Trp261-Ser262), leaving the terminal Trp261 residue in the substrate binding pocket, thus inactivating the CP.

The surface near the cleft in-house some hydrophobic residues which acts as site for binding of the cdE2 glycoprotein and is referred to as hydrophobic pocket. Structure comparison with other known alphavirus CP shows that the loop region found in the hydrophobic pocket is flexible and of variable length. The protein-protein interactions between cdE2 and CP are crucial for alphavirus budding and thus life cycle of virus. The Pro404 of the H-L-H motif of cdE2, fits in the hydrophobic pocket of CP. The Pro404 of cdE2 makes hydrophobic contacts with the residues of the CP hydrophobic pocket. Studies in past have reported the various techniques to design peptidomimetic molecules against viral proteases (Dhindwal et al., 2016; Randolph and DeGoey, 2004; Smith Iii et al., 1994; Choi et al., 2009; 2012). Such techniques can be employed to target the cdE2-CP interactions and design mimetics against the hydrophobic pocket.

Literature shows that targeting hydrophobic pocket through dioxane and picolinic acid (PCA) structurally mimics the binding of pyrrolidine ring of proline residue and is predicted to disrupt the cdE2-CP molecular interactions and indirectly inhibit the virus budding and replication process. Small heterocyclic compounds (MDA and EAB) were selected based on the structure similarity to PCA. Results of molecular docking of MDA and EAB suggested that these two molecules have good binding affinity for the hydrophobic pocket of CVCP. Further, the molecular interactions were analyzed biophysically through SPR and fluorescence experiments. The values of the dissociation constants (K_D) obtained from SPR for binding of MDA and EAB

to CVCP were 1.2×10^{-3} M and 0.2×10^{-9} M, respectively. Both MDA and EAB quench the CVCP fluorescence and exhibit red shift. Since the cdE2-CP interaction takes place through the Pro404 residue lying in the loop region of the H-L-H motif of cdE2. Additionally, Pro404 of cdE2 and the residues lining the hydrophobic pockets of the CVCP are conserved. Thus, the derivatives of MDA and EAB molecules designed and developed based on the complimentary shape and chemistry of the hydrophobic pocket are expected to have broad spectrum antiviral activity against alphaviruses.

In addition, structure based linear B-cell epitopes have been identified on the surface of CVCP protease domain using *in silico* techniques. The regions 148-160 and 170-190 were identified as the probable epitopic regions involved in causing antigenic reaction in the host. Previous studies have demonstrated that a series of overlapping epitopes in the C-terminal region (140-210) of CHIKV capsid were recognized by monoclonal antibodies and intact C-terminus is required for antigenic structure of CP (Goh et al., 2015). Additionally, the region 148-160 lying in sub-domain 1 includes the NES region of the capsid protein whereas region 170-190 is part of the flexible linker-loop region. The latter also lies in close proximity to the hydrophobic pocket of the CP. Since targeting mimotopes can trigger autoimmune reaction, the identified epitopes were mapped to the human proteome to evaluate for presence of mimotopes. No significant hits were obtained which suggest that these identified epitopes could potential be used for developing molecular diagnostics and therapeutics. The study elucidates the importance of hydrophobic pocket of chikungunya CP and its involvement in interaction with viral cdE2 protein. This study adds to the present understanding of chikungunya virus capsid protein structure and function, which further helps in development of structure based antivirals against alphaviral diseases.

Chapter 4

Characterization of trans-proteolytic activity of chikungunya virus capsid protease

4.1 Abstract

Alphavirus capsid protein (CP) is a multifunctional protein. It comprises of two domains, the RNA binding N-terminal domain and the C-terminal protease domain. The N-terminal domain, which is highly disordered is involved in binding to the genomic RNA, PPIs (protein-protein interactions) that lead to the CP dimerization and the inhibition of host cell transcription. The CP protease domain possesses a chymotrypsin-like fold with a conserved catalytic triad (Ser, Asp and His residues). The C-terminal CP has autoproteolytic activity and plays a critical role in initiation of the structural polyprotein processing and the viral life cycle

This chapter reports the proteolytic activity of CHIKV CP in *trans* and development of high-throughput screening (HTS) assay for inhibitors screening. CHIKV CP is a potential target because of its important role in structural polyprotein processing, contain signals for cytoplasmic trafficking (NES) and nucleocapsid core formation inside host cells. A fluorescence resonance energy transfer (FRET) based assay for screening of CHIKV capsid protease inhibitors has been developed. This assay involves the use of fluorogenic peptide substrates for developing HTS assay. The fluorogenic substrate peptide was derived from the capsid protease cleavage site flanked by fluorophore and quencher at both the termini. When the capsid protease cleavage site incorporated in the fluorogenic peptide substrate is cleaved by active capsid protease, the FRET decreases and the fluorescence emission intensity for the donor fluorophore increases.

Kinetic parameters using fluorogenic peptide substrates for the chikungunya virus capsid protease were estimated. The K_m and K_{cat}/K_m values were calculated and found to be $0.754 \pm 0.11 \mu\text{M}$ and $1.72 \times 10^5 \text{ M}^{-1} \text{ min}^{-1}$, respectively. The effect of serine protease inhibitors, N-p-Tosyl-L-phenylalanine chloromethyl ketone (TPCK) and Benzamidine hydrochloride (BHC) on CVCP protease activity has also been determined. The value of dissociation constant (K_i) obtained for BHC and TPCK were $115 \mu\text{M}$ and $220 \mu\text{M}$, respectively. In conclusion, this chapter explains the development of a HTP method for testing and screening inhibitors against the proteolytic activity of CHIKV capsid protease. The basic molecular architecture of the active site and the catalytic residues are highly conserved among the members of serine protease family including alphaviruses CPs. Therefore, this assay can be used for development of anti-alphaviral inhibitor screening.

4.2 Introduction

Chikungunya virus (CHIKV), a well-known high threat virus that account millions of deaths every year. Chikungunya fever is a re-emerging viral disease, rapidly increasing in geographic range and frequency (Abdelnabi et al., 2017; Cavrini et al., 2009; Morrison et al., 2014). The chikungunya virus is transmitted by the bite of female mosquitoes, *Aedes albopictus* and *Aedes aegypti*. In humans, CHIKV infection is associated with an array of complications from mild fever, rashes, musculoskeletal pain to persistent arthralgia (Larrieu et al., 2010). The CHIKV associated clinical manifestations includes severe dermatological lesions, lymphopenia, encephalitis and neonatal encephalopathy (Sourisseau et al., 2007). The disease was first reported in 1953, from the serum of a febrile patient in Tanzania (Ross et al., 1956) and afterwards numerous outbreaks occurred in various parts of the world. Later, major outbreak in Kenya had spread the disease in Comoros, Madagascar, La Reunion Island, West Africa, South East Asia and Europe and (Sergon et al., 2005; 2008; Rezza et al., 2007). CHIKV outbreak in La Reunion Island in 2005, had affected one third of its population (Sergon et al., 2005). In 2006, CHIKV epidemic in India, had affected more than 1.4 million people (Shah et al., 1964; Kumar et al., 2010; Padbidri et al., 1978; Mavalankar et al., 2008). In 2008, National Institute of Allergy and Infectious diseases (NIAID) agency of U.S has listed CHIKV as category C pathogen (Schwartz et al., 2010). The recurrence of chikungunya fever due to mutation in E1 glycoprotein (Ala226Val) (Schuffenecker et al., 2006) and increased pathogenicity from arthralgia to encephalitis and meningoencephalitis has made CHIKV a noteworthy pathogen that demands immediate attention. Presently, there is no specific medicine or vaccine available against CHIKV infection that compels the study of alphaviral biology and development of anti-CHIKV drugs.

Alphaviruses are enveloped viruses and consist of single-stranded positive sense RNA transmitted by mosquitoes. Alphavirus genus include viruses like Sindbis virus (SINV), Semiliki Forest virus (SFV), Venezuelan Equine Encephalitis virus (VEEV), Ross River virus (RRV) etc. Members of this genus are classified into New World and Old World viruses on the basis of the mechanism they employed for host transcription shut off, disease presentation and mortality rate. Old World viruses include SINV, CHIKV and SFV etc. employ the nsp2 protein to shut host transcription off, arthralgic in disease presentation and have low mortality rate. New world viruses include WEEV, VEEV etc. utilize their capsid protein to shut off the host cell transcription, cause encephalitis and have high mortality rate (Garmashova et al., 2007; Weaver et al., 2012; Powers et al., 2001). Alphavirus genus contains 29 members affecting a range of hosts such as humans, fish, rodents, horse etc. Corresponding to the other alphavirus members, CHIKV genome size is approximately ~ 11.7 kb, capped at 5' end and poly-adenylated at 3' end

(Baron et al., 1996; Vasiljeva et al., 2000). About 240 copies of capsid protein encapsidate the virus RNA genome forming nucleocapsid core (NC). The diameter of the mature alphavirus virion is 70 nm with host derived envelope embedded with 80 spikes in T=4 icosahedral symmetry. Each spike is made up of trimers of surface glycoproteins (E1 and E2) heterodimers (Mukhopadhyay et al., 2006; Soonsawad et al., 2010; Strauss et al., 1994).

The N-terminal two-third section of alphavirus genome encodes for non-structural proteins and C-terminal one third section of genome encodes structural proteins. The translation of genomic 49S RNA produce non-structural polyprotein (Weaver et al., 2012; Powers et al., 2001). The genomic 49S RNA via minus strand RNA intermediate also transcribes the 26S sub-genomic RNA that translates in to structural polyprotein. The structural polyprotein comprises of CP, E3, E2, 6K and E1 proteins. The capsid protein (CP) is a multifunctional protein. It is involved in autoproteolysis, capsid–glycoprotein interactions, nucleocapsid core formation, host transcription down regulation and contain signals for nuclear and cytoplasmic trafficking (Choi et al., 1991; Melancon et al., 1987). CP consists of two domains, the RNA binding N-terminal domain and the C-terminal protease domain. The N-terminal domain is highly disoriented and rich in basic residues, is responsible for RNA genome encapsidation, host transcription down regulation and capsid-capsid dimerization (Lulla et al., 2013; Owen et al., 1996). The C-terminal domain possess chymotrypsin like serine protease and is highly conserved. The C-terminal protease domain of CP has *cis*-proteolytic activity and stated to not possess trans-activity (Choi et al., Hahn et al., 1990). The alphavirus CP cleaves itself at W/S scissile bond at the C-terminal end and after the cleavage, the terminal Tryptophan (W261 in CVCP) interacts with the residues of catalytic triad through free carboxylic group and inhibits the further protease activity (Aggarwal et al., 2012). Conversely, other studies on Semliki Forest virus capsid protein by Morillas et al., showed that the truncation of 7 residues from C-terminus reinstate enzymatic activity (Morillas et al., 2008). Recent studies on Aura virus capsid protein (AVCPΔ2) have demonstrated the FRET (Fluorescence resonance energy transfer) based trans-protease activity after the truncation of two residues from C-terminus (Aggarwal et al., 2014). The structural analysis of AVCPΔ2 crystal structure further gave the structural insights into catalytic architecture and folding of whole protease domain. Similar to members of serine protease family, capsid protease catalytic triad (His139, Asp161 and Ser213 in CVCP) and active site molecular architecture in alphavirus is highly conserved (Aggarwal et al., 2014; Choi et al., 1996; 1997). Moreover, the GDSG motif having active site serine nucleophile is also conserved among chymotrypsin like serine protease and alphavirus family.

The protease activity of alphavirus CP, including CVCP is very critical for structural polyprotein processing and viral life cycle. Because of its key role in alphaviruses, it is considered as an important drug target. Fluorogenic assay based reporter system for characterization of protease activity and subsequent inhibitors screening was extensively used for viral proteases like 3C proteases, HCV (hepatitis C virus) NS3/4A protease, Human immunodeficiency virus (HIV), West Nile virus serine protease and bacterial SUMO proteases, Dengue NS2B-NS3 protease (Blanchard et al., 2004; Mueller et al., 2008; Sudo et al., 2005; Yang et al., 2013). Because of simplicity, time saving and highly sensitivity of this technique, similar strategy has been used for CVCP protease activity. Fluorogenic peptide substrate containing DABCYL [4-(4-dimethylaminophenyl-azo)benzoic acid] and EDANS (5-[(2-aminoethyl) amino]naphthalene-1-sulfonic acid) on the N and C terminus of CVCP cleavage site was used for monitoring CVCP protease activity. In this FRET pair DABCYL being a quencher, quenches the fluorescence of EDANS (fluorophore) placed at its certain proximity. After the cleavage of fluorogenic peptide substrate, the peptide gets cleaved into two segments separating DABCYL and EDANS. This cleavage leads to increase in the intensity of fluorescence and reduction of FRET signal. This method can be used for high-through-put screening of protease inhibitors as in the presence of inhibitor, substrate peptide will not get cleaved and thus no increase in fluorescence. To estimate the quality of protease assay, the Z' factor was determined as per method reported by Zhang et al., 1999. The Z' factor was determined to evaluate the reproducibility and reliability of FRET assay. The fluorogenic peptide substrate with FRET pair flanking N and C termini of protease cleavage site was used for determine the Z' factor. The fluorescence readings were recorded for the assay reactions containing substrate peptide, both in the presence and absence of enzyme. The Z' factor can be defined as:

$$Z' \text{ factor} = 1 - \frac{3(\sigma_p + \sigma_n)}{|\mu_p - \mu_n|}$$

where μ_n , μ_p , σ_n and σ_p are the standard deviations; and p and negative n are averages (μ) of positive and negative controls, respectively. A screening assay with value of Z' factor greater than 0.5 can be considered as an efficient and authentic screening assay.

Serine proteases are the most widely studied group of proteases because of their role in pathological and physiological processes in host. The catalytic mechanism of serine protease is shown in Fig. 4.2.1. Upon binding of substrate to serine protease, the OH group of nucleophilic serine of catalytic triad attacks on the carbonyl carbon of the substrate followed by acceptance

of hydrogen by His. These series of reactions lead to the formation of tetrahedral intermediate. The transfer of electron pair from scissile bond of substrate peptide to H of His results in scissile bond cleavage. Then the electrons from the carbonyl oxygen move back to form the double bond again and results in acyl enzyme intermediate. The hydrolytic reaction of acylated serine through water in presence of His and Asp helps in enzyme regeneration.

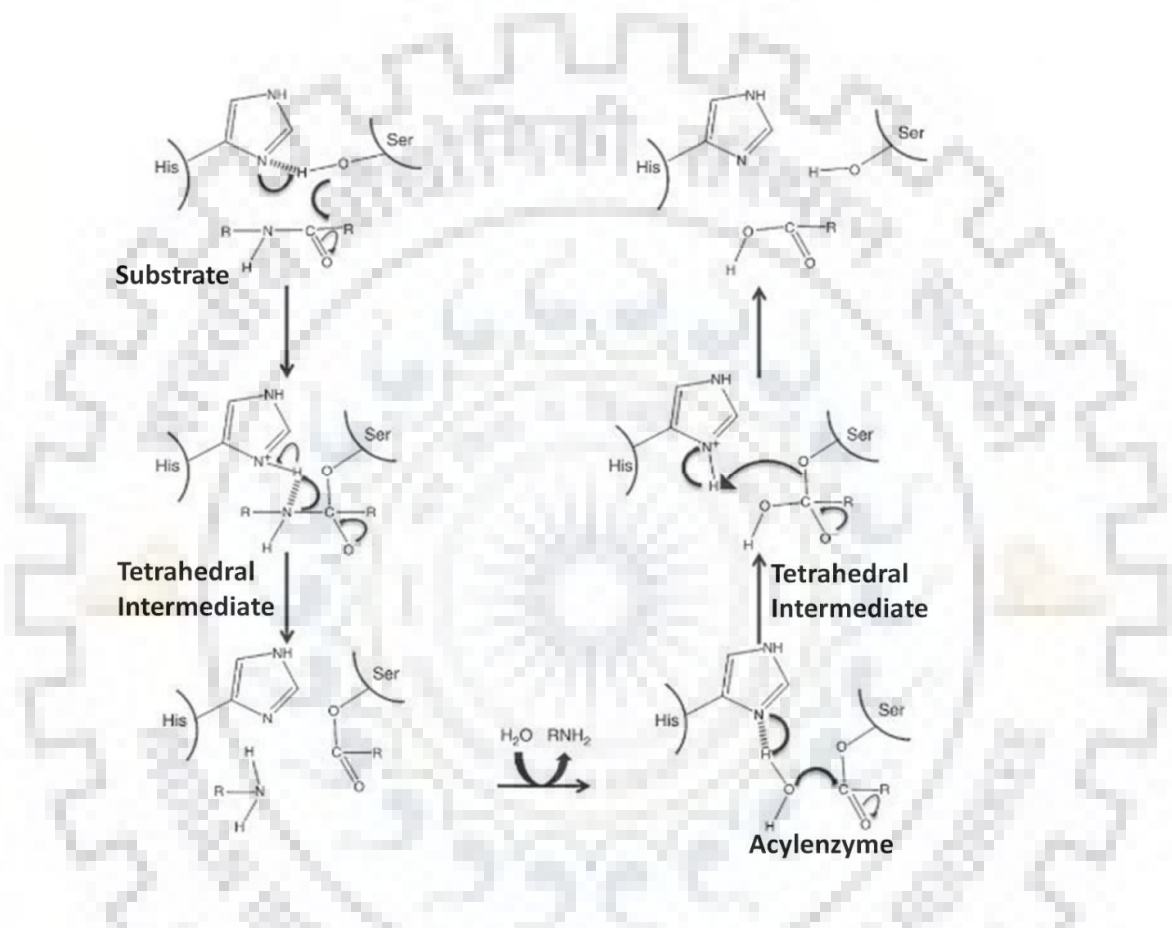


Figure 4.2.1: Catalytic mechanism of serine protease. The binding of the substrate to the serine protease induces the nucleophilic attack on carbonyl carbon of substrate by OH group of serine residue of catalytic triad. This is followed by acceptance of hydrogen by N-atom of His residue of catalytic triad and subsequently formation of tetrahedral intermediate. The shifting of scissile bond electrons of substrate peptide to H of His is responsible for substrate peptide bond cleavage. The shifting of double bond electrons back to carbonyl oxygen results in formation of acyl enzyme intermediate which upon hydrolytic reaction helps in enzyme regeneration (Tomar et al., 2017).

Several serine protease inhibitors are known which can be used for therapeutic intervention for varied disease conditions. Serine protease inhibitors include peptide inhibitors or chemical inhibitors that can bind reversibly or irreversibly to the proteases (Powers et al., 2002). The effect of serine protease inhibitors and BHC and TPCK on CVCP protease activity has also been determined. BHC and TPCK are reversible and irreversible protease inhibitors,

respectively. BHC is a reversible inhibitor of trypsin, trypsin-like proteases and serine proteases. Benzamidine binds to the S1 specificity pocket of proteases and inhibits the protease activity competitively (Renatus et al., 1998; Wu et al., 2007). The amidino group of benzamidine forms salt bridge with carboxylate group of Asp189. Additionally, the contact of OH group of Ser190 with Gly214 carbonyl group through a water molecule is further associated with stabilization of salt bridge (Yamane et al., 2011; Renatus et al., 1998; Bode et al., 1990) (Fig. 4.2.2).

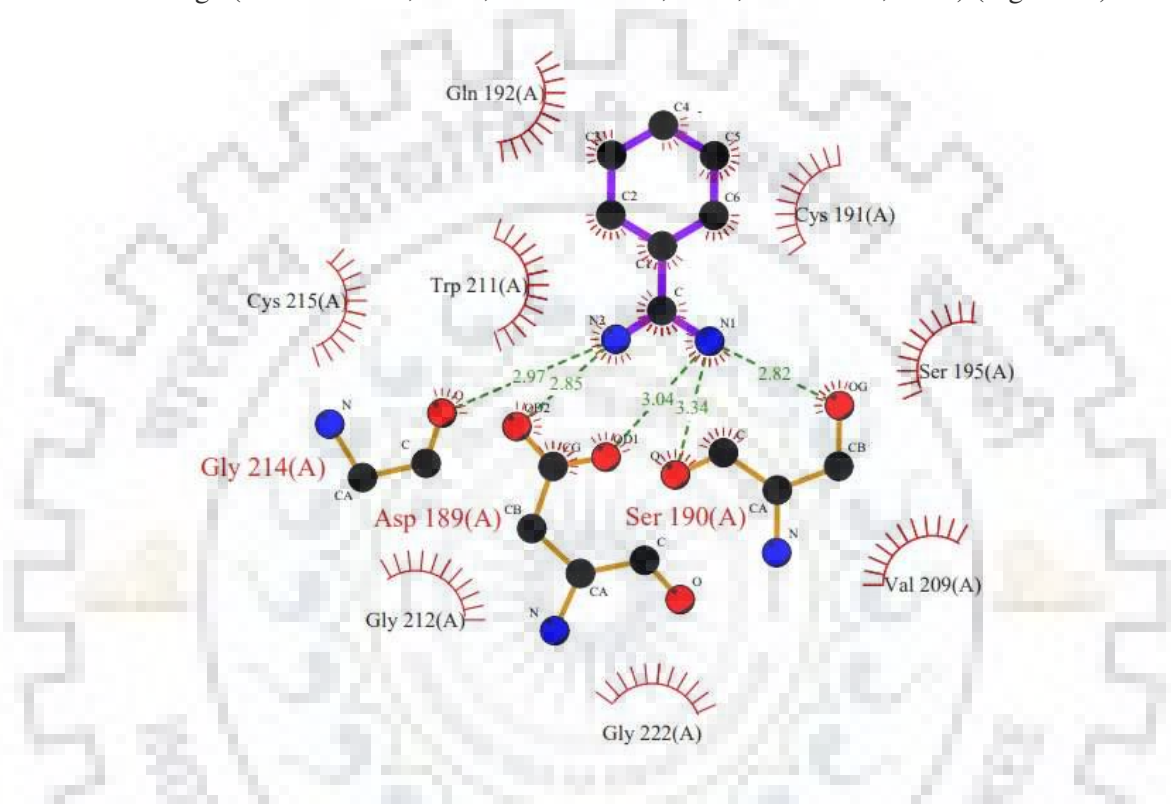


Figure 4.2.2: The LigPlot analysis of benzamidine binding to protease S1 specificity pocket, PDB ID: 3ATL (Yamane et al., 2011).

TPCK (chymotrypsin substrate derivative) is an irreversible serine and cysteine protease inhibitor associated with treatment of hypoxic-ischemic brain injury and post traumatic brain injury in rat models (Polgar et al., 1989; Feng et al., 2003; Powers et al., 2002). Chloromethyl ketone inhibitors (TPCK) induces irreversible alkylation of active site His residue of serine proteases and inhibits protease activity (Fig. 4.2.3). Several studies have reported the chloromethyl ketone inhibitors act as transition-state inhibitors. The interaction of active site Ser195 with carbonyl group of inhibitor results in the formation of tetrahedral adduct which facilitates the irreversible alkylation of active site His residue. The methylene carbon of these inhibitors forms two covalent bonds with active site residues, one with His64 (Otto et al., 1997)

and other with Ser221 (Crawford et al., 1988). The mechanism of serine and cysteine protease inactivation by halomethyl ketone inhibitors (where X= F, Cl, Br and I) is shown in Figure 4.2.3.

In conclusion, this chapter explains the development of a HTS method for testing and screening of inhibitors against the proteolytic activity of CHIKV capsid protease.

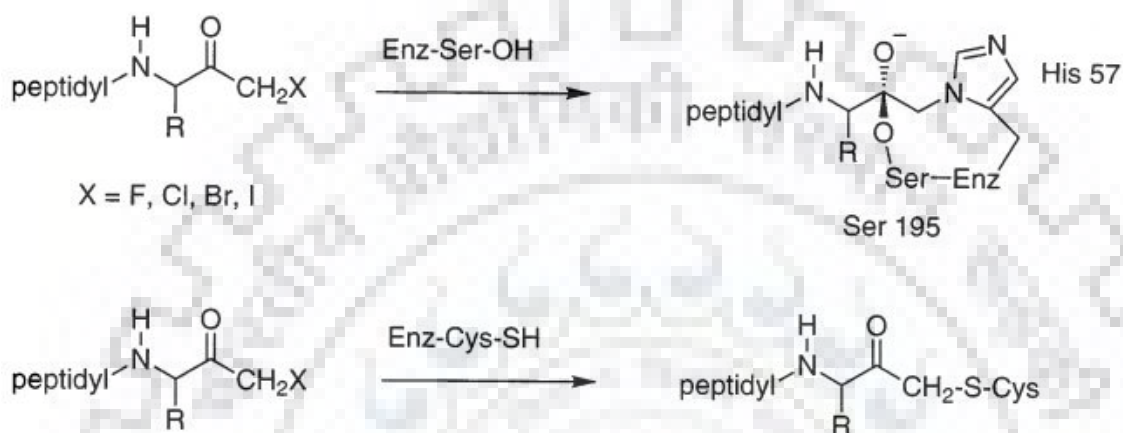


Figure 4.2.3: Mechanism of inactivation of serine and cysteine proteases by halomethyl ketone inhibitors (Powers et al., 2002; Walker et al., 2001).

4.3 Materials and Methods

4.3.1 Expression of active CVCP

Chemically competent cells of *Escherichia coli* strain *Rosetta (DE3)* were transformed with recombinant plasmid (pET28c) containing N-terminal His-tagged CVCP capsid protein gene (residues 106-259). The expression conditions for active CVCP were optimized to produce the protein in soluble form. The LB agar plate was used for plating the transformed cells in presence of kanamycin (50 µg/ml) and chloramphenicol (35 µg/ml), incubated at 37 °C overnight. Fresh autoclaved culture tube containing LB broth was inoculated by picking up a colony from the plate, followed by incubation at 37 °C overnight with constant stirring. The overnight grown culture was used as inoculum to raise 1L secondary culture. The secondary culture was grown at 37 °C till the OD reaches to 0.7 (OD₆₀₀), afterwards the induction was done using 0.4 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). After induction with IPTG, culture was grown at 18 °C overnight with constant agitation. The induced active CVCP cells were collected as pellet by centrifugation at 8,000 x g for 15 min at 4 °C and analyzed for expression and solubility using 15 % SDS-PAGE.

4.3.2 Purification of active CVCP

Frozen pellets of active CVCP were thawed and re-suspended in 25 ml of lysis buffer (50 mM Tris-HCl pH 7.6, 10 mM imidazole and 100 mM NaCl) for the purification of active CVCP. This suspension of cells were disrupted using the French Press (Constant Systems Ltd, Daventry, England) followed by centrifugation at 16,000 x g for 45 min at 4 °C to separate the supernatant and pellet. The soluble protein was purified by using affinity chromatography by using Ni-NTA agarose beads pre-equilibrated with the binding buffer (50 mM Tris-HCl pH 7.6, 100 mM NaCl). The supernatant was incubated with the Ni-NTA agarose beads for 45 min at 4 °C. Successive three washes were made with varied concentrations of Imidazole (ranges from 30 mM to 100 mM) along with the binding buffer. The protein was finally eluted by using 250 mM Imidazole. The pooled eluted fractions were incubated with TEV protease for cleavage of His-tag during dialysis. The cleaved His-tag, were removed from the protein samples by reloading it onto Ni-NTA column. All the eluted fractions were analyzed by 15 % SDS-PAGE. The purified protein was concentrated by a 3 kDa cutoff Amicon (Millipore) to ~5 mg/ml. The protein concentration was estimated by UV absorbance spectroscopy at a wavelength of 280 nm.

4.3.3 CVCP *trans*-protease activity

FRET based proteolytic activity assay was established for CVCP by utilizing DABCYL and EDANS as FRET pair, located at the N and C terminal, respectively to the synthetic fluorogenic peptide. In this FRET pair, EDANS has the fluorescence property (fluorophore) and DABCYL quenches the fluorescence produced by EDANS. The synthetic peptide having FRET pair flanking the termini of CVCP protease cleave site, was acquired from Biolinkk (New Delhi, India). The substrate peptide sequence was taken from CVCP protease cleavage site encompassing W/S scissile bond. On incubation of fluorogenic peptide substrate with CVCP protease, the peptide gets cleaved into two fragments separating DABCYL and EDANS. This cleavage of substrate peptide results in reduction of FRET signal and increase in the fluorescence emission intensity for the donor fluorophore (Fig. 4.3.3.1).

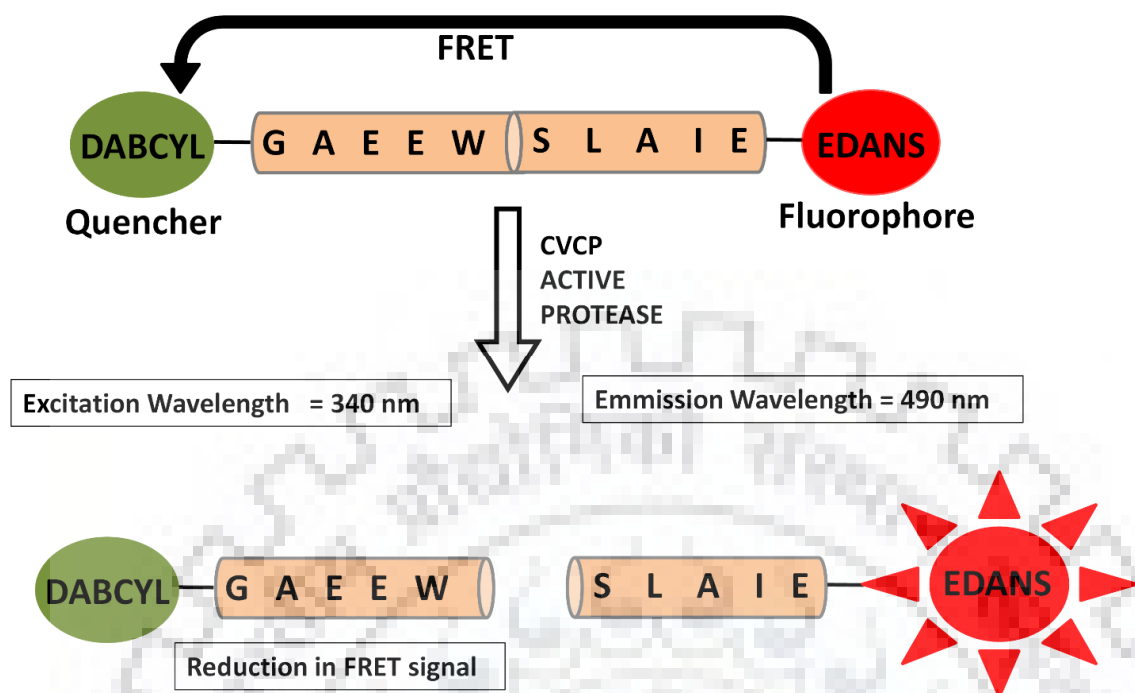


Figure 4.3.3.1: Fluorescence resonance energy transfer (FRET) protease assay. FRET based protease assay for active CVCP has been developed. The assay involves the use of fluorogenic peptide substrate containing sequence from the C-terminal end of capsid protease cleavage site that includes the scissile bond W/S residues. The quencher (DABCYL) and fluorophore (EDANS) are attached at the N and C terminus of the peptide substrate, respectively. Upon incubation with active CVCP, the fluorogenic peptide substrate containing the cleavage site is cleaved by active capsid protease, the FRET decreases and the fluorescence emission intensity for the donor fluorophore increases. This assay can be used for screening inhibitors against the CHIKV capsid protease activity as in presence of inhibitors, increase in the fluorescence will not be observed.

The stock solution of fluorogenic substrate peptide was made in 100 % DMSO. The final concentration of Dimethyl sulfoxide (DMSO) in reaction mixture was kept below 1 % (v/v). The buffer was optimized for the CVCP protease activity on incubating it with the substrate peptide. The CVCP protease assay was carried out in 20 mM HEPES buffer (pH 7.0). The total volume of the reaction mixture used was 100 μ l and the reaction was set up in 96 well plate (Corning, USA, Cat. No. 3686). Fluorescence microplate reader, Cytation 3 (BioTek Instruments, Inc) was used to record the increase in fluorescence intensity at regular intervals of time. The excitation and the emission wavelengths for the fluorogenic peptide substrate was 340 nm and 490 nm, respectively. Kinetic parameters using a range of fluorogenic peptide substrates concentrations (0.4 to 2.5 μ M) for the chikungunya virus capsid protease activity were estimated. The concentration of both the enzyme (CVCP) as well as substrate (fluorogenic peptide) was optimized for CVCP activity analysis. The initial velocity (V_i) of enzyme at range of substrate concentrations was calculated. The kinetic parameters were estimated by Michalies-Menten as well as Lineweaver Burk Plots. The readings obtained as fluorescence units were normalized

with control readings before doing activity analysis. Additionally, Z' factor for CVCP protease assay was calculated by using fluorogenic peptide substrate with FRET pair at N and C termini of CVCP protease cleavage site. The reaction mixture including the substrate peptide was used for recording the fluorescence in the presence and absence of CVCP enzyme. Chymotrypsin and CVCP inactive protein were used as positive and negative controls.

4.3.4 Effect of inhibitors on CVCP protease activity

FRET based protease assay was previously used for protease activity reporter system and subsequent inhibitors screening for viral proteases like 3C proteases, HCV NS3/4A protease and bacterial SUMO proteases (Yang et al., 2013; Blanchard et al., 2004; Sudo et al., 2005). In the present study, FRET based protease assay was used for CVCP *trans*-protease activity. The FRET assay was also used for testing and screening of inhibitors against the proteolytic activity of CHIKV capsid protease. In native CVCP (inactive) after the *cis*-proteolytic activity, the C-terminal Trp residue left bound to the S1 specificity pocket and inhibits further catalytic activity. The removal of terminal tryptophan from the CVCP retains the protease activity. Based on this, the known serine protease inhibitors were tested for inhibition of CVCP protease activity. The effect of serine protease inhibitors TPCK and BHC on CVCP protease activity has been determined. The CVCP protease was incubated with inhibitors for 0.5 h tailed by addition of substrate in the reaction mixture. The different concentrations of inhibitors tested were 0 μ M, 40 μ M and 50 μ M and final volume of the reaction mixture was 100 μ l. The protocol used for study the FRET based inhibition assay was same as described above.

4.4 Results and Discussion

4.4.1 CVCP Expression and Purification (active)

Alphavirus capsid protein has *cis*-proteolytic activity that cleaves itself at W/S once in a life cycle. From the C-terminal end of CVCP, two residues were removed counting C-terminal W261. The truncation of C-terminal tryptophan would allow the substrate to access the catalytic site and consequently leads to restoration of protease activity that cleaves substrate at W/S cleavage site (Fig. 4.4.1.1).

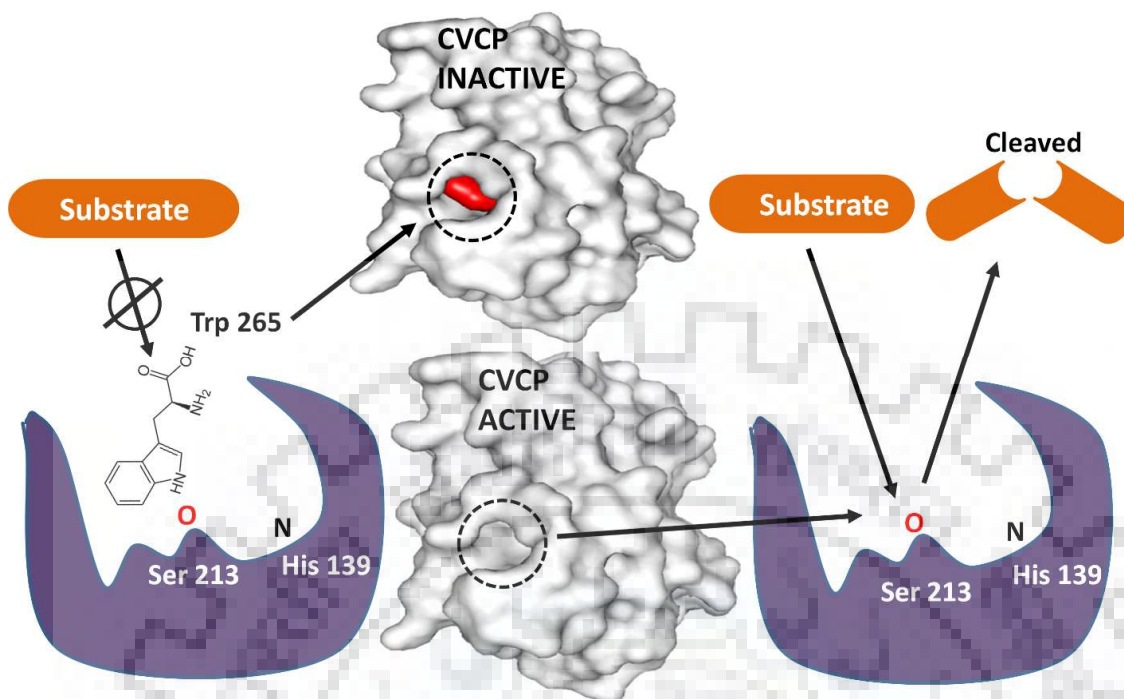


Figure 4.4.1.1: Schematic representation of alphavirus active and inactive protease. After the *cis*-proteolytic activity of CVCP protease domain, the C-terminal Trp267 remains bound to catalytic site and inhibits further protease activity (left). In the absence of Trp267, the protein acquires its catalytic property as the substrate can easily access the active site (right).

Recombinant plasmid (pET28c-CVCP Δ 2) was transformed into *Rosetta* (DE3) cells for production of CVCP active protein (106-259 residues). Expression conditions for CVCP active constructs were optimized to get the soluble protein in good quantity. Both the induction temperature and time was optimized. The IPTG (0.4 mM) was used for expression of CVCP protein. The IPTG induced culture was incubated overnight at 18 °C followed by centrifugation. The harvested cells were stored at – 80 °C till further processing. The CVCP active cells were analyzed for the expression of soluble protein using 15 % SDS-PAGE (Fig. 4.4.1.2). After optimization of CVCP active protein expression, purification was done using Ni-NTA affinity chromatography. The purified CVCP protein displayed a single band, equivalent to ~17 kDa, in ~12 % SDS-PAGE (Fig. 4.4.1.3). The purified protein was concentrated to ~5 mg/ml and subjected to crystallization trials using different Hampton screens. The vapor diffusion method was used for crystallization of CVCP protein. Initial crystallization trials gave promising hits as polycrystalline aggregates and microcrystal have been observed in Hampton index screens. The obtained crystallization conditions will be used for optimization of CVCP crystallization using both sitting drop and hanging drop methods.

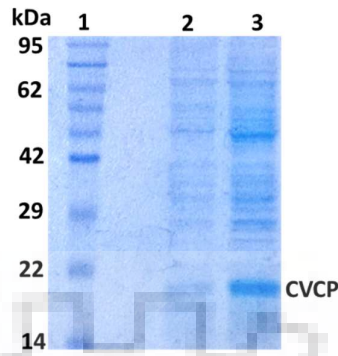


Figure 4.4.1.2: Expression of active CVCP: SDS-PAGE analysis of expression and solubility of recombinant active chikungunya (110-259) capsid protein Lane 1; molecular-weight markers (kDa), Lane 2; Supernatant, Lane 3; pellet

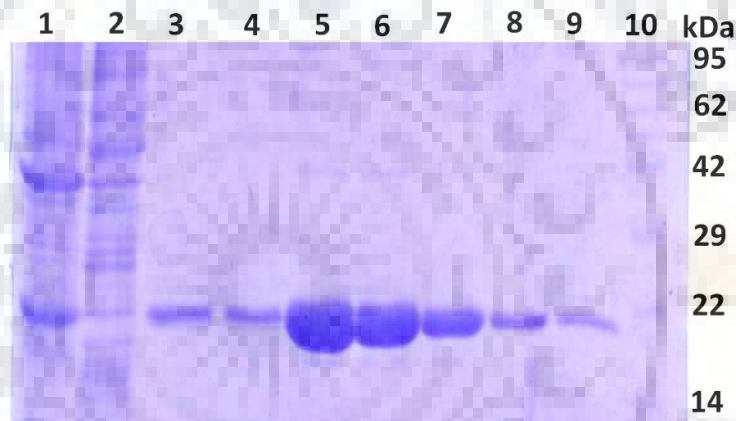
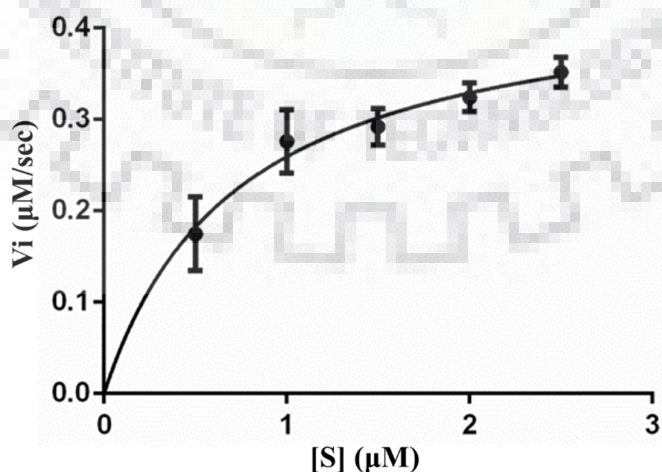


Figure 4.4.1.3: Purification of active CVCP: SDS-PAGE and active CVCP shows the protein purified to homogeneity. Lane 1; pellet), lane 2; supernatant containing soluble protein fraction, Lane 3; Wash I, Lane 4; Wash II, Lane 5-9; Elutions, Lane 10; molecular-weight markers (kDa)

4.4.2 CVCP *trans*-protease activity

Truncated CVCP protein (CVCP Δ 2) was used for FRET based *trans*-protease assay. The substrate peptide sequence was derived from C-terminus CVCP and N-terminus E3 glycoprotein including W/S scissile bond. The substrate fluorogenic peptide containing CVCP cleavage site (W/S) is flanked by EDANS (fluorophore) and DABCYL (quencher) at opposite ends. All the fluorogenic protease assay reactions were set up in 20 mM HEPES buffer (pH 7.0) with the final reaction volume of 100 μ l. The optimal excitation and emission wavelength of EDANS is 340 nm and 490 nm, respectively. The CVCP active enzyme on incubation with substrate fluorogenic peptide results in increase in fluorescence, previously quenched by presence of DABCYL in proximity of EDANS. The increase in fluorescence signal was recorded at regular time intervals.

The readings obtained as fluorescence units were normalized with control readings (reaction having no enzyme). The final fluorescence readings termed as RFU's (Relative Fluorescence Units) got after subtracting the control readings from fluorescence readings of the enzyme reaction. The initial velocity (V_i) of enzyme at range of substrate concentrations was determined. Kinetic parameters using fluorogenic peptide substrates for the chikungunya virus capsid protease were estimated. The rate of the CVCP protease reaction has increased with increasing substrate (fluorogenic peptide) concentration and thus followed the Michaelis–Menten kinetics. A double reciprocal Lineweaver–Burke plot of $1/V_i$ against $1/[S]$ was also plotted to determine the values of K_m and K_{cat}/K_m . The K_m and K_{cat}/K_m values were found to be $0.754 \pm 0.11 \mu\text{M}$ and $1.72 \times 10^5 \text{ M}^{-1} \text{ min}^{-1}$, respectively (Fig. 4.4.2.1). Both positive and negative control reactions were also set up to validate the CVCP protease assay. Chymotrypsin and inactive CVCP was used in as a positive and negative control, respectively. As expected, chymotrypsin showed increase in fluorescence signal on addition of substrate fluorogenic peptide into the reaction. On the other hand, inactive CVCP showed decrease in fluorescence signal in the same reaction. The concentration of fluorogenic peptide substrates used for fluorescence measurements was very low $\sim 1 \mu\text{M}$ and thus inner filter effect was insignificant. Inner filter effects are responsible for nonlinear curves and distorted emission spectra. In CVCP protease assay, fluorescence intensity was linearly proportional to the concentration of the substrate peptide used. The value of Z' factor calculated for CVCP protease assay was 0.64, which reflects high signal to noise ratio. It confirms the sensitivity and reproducibility of the protease assay. The CVCP protease assay can be used for high-throughput screening (HTS) of inhibitors against the proteolytic activity of CHIKV capsid protease.



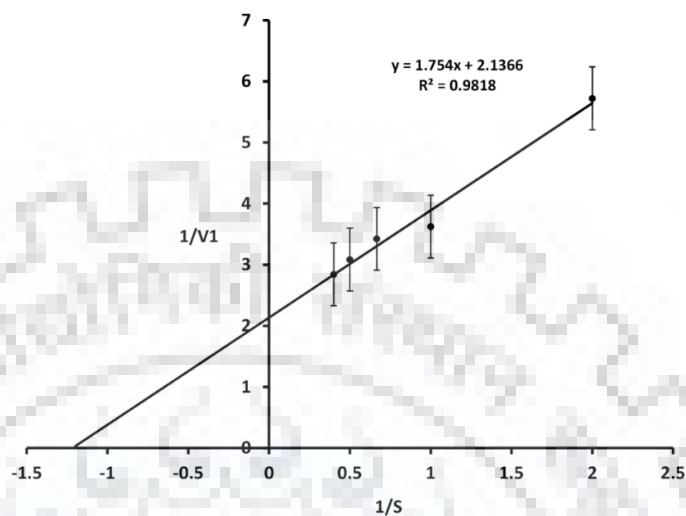


Figure 4.4.2.1: Kinetic analysis of active CVCP: (Upper panel) Kinetic parameters for the chikungunya virus capsid protease using fluorogenic peptide substrates were estimated. The different concentrations of the substrate used range from 0.4 μM to 2.5 μM . The initial reaction rates were plotted against different substrate concentrations. GraphPad Prism 6.0, software was used to fit the kinetic data using Michaelis-Menten equation. (Lower panel) The lineweaver-burk plot of $1/V_i$ against $1/[S]$ (double reciprocal of V_i and S) was generated. The intercept and the slope were calculated according to the equation $y = mx + c$. The values of V_{max} and K_m were determined from the lineweaver-burk plot equation and the K_{cat} was also calculated by dividing V_{max} with the enzyme concentration.

4.4.3 Effect of inhibitors on CVCP protease activity

The serine protease inhibitors TPCK and BHC were tested against CVCP protease activity, to unveil the inhibition, if any. From the literature review, it has been found that the BHC is a reversible inhibitor of trypsin, trypsin-like proteases and serine proteases and has a reversible mode of binding. While, the TPCK being a serine protease inhibitor has an irreversible mode of binding. CVCP protease inhibition assay was carried out in 100 μl reaction volume containing 20 mM HEPES buffer (pH 7.0) using 96-well fluorescence plate. The CVCP protease was incubated with the inhibitors for 0.5 h followed by the addition of fluorogenic substrate peptide. Then the fluorescence reading were acquired at different time intervals. For each inhibitor, the different concentration of inhibitors tested were 0 μM , 40 μM and 50 μM . In the presence of these inhibitors, the intensity of fluorescence signal was reduced as compared to the control reaction in which no inhibitor was added. For the determination of inhibitor constant (K_i) initial velocity (V_i) of the active CVCP (enzyme) at each concentration of tested compound was

determined. The obtained values were normalized by subtracting the reaction readings with the reaction having no enzyme. A Lineweaver-Burk plot was generated to determine the values of inhibition constant K_i . The value of inhibition constant (K_i) obtained for BHC and TPCK were 115 μM and 220 μM , respectively (Fig. 4.4.3.1 and 4.4.3.2). In the steady state kinetics, BHC and TPCK gave competitive and uncompetitive pattern, respectively (Fig. 4.4.3.1 and 4.4.3.2). This assay development has paved the way for the development of a HTS method for testing and screening inhibitors against the proteolytic activity of CHIKV capsid protease. Like other serine proteases, the basic molecular architecture of the active site and the residues involved in catalysis are highly conserved in alphavirus CPs. Therefore, this assay can be used for development of anti-alphaviral inhibitor screening.

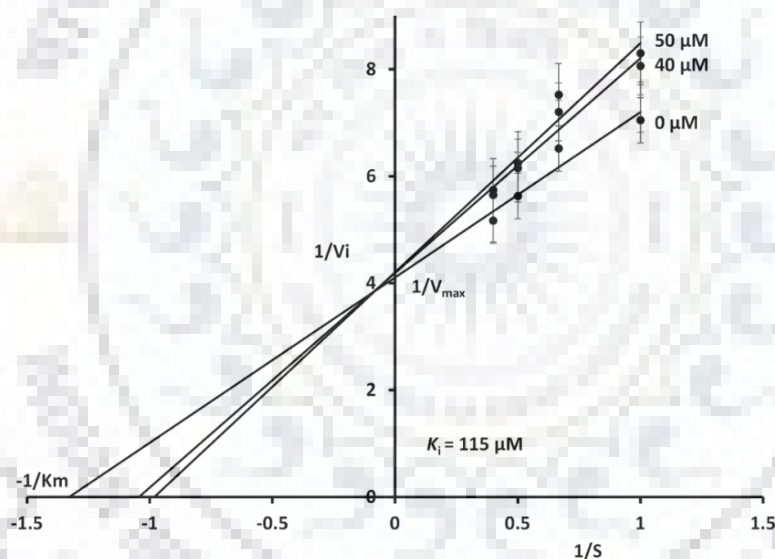


Figure 4.4.3.1: Determination of Inhibition constant for Benzamidine hydrochloride (BHC): The lineweaver-burk plot of $1/V_i$ against $1/[S]$ (double reciprocal of V_i and S) was generated. The slope of the plot gives K_m/V_{max} , x-intercept gives $1/V_{\text{max}}$ and y-intercept gives $-1/K_m$. As seen from the plot, with the increase in the inhibitor concentration the value of apparent K_m increases (y-intercept moves right) V_{max} remains constant. The BHC showed competitive mode of inhibition for CVCP protease activity. The value of inhibition constant (K_i) obtained for BHC was 115 μM .

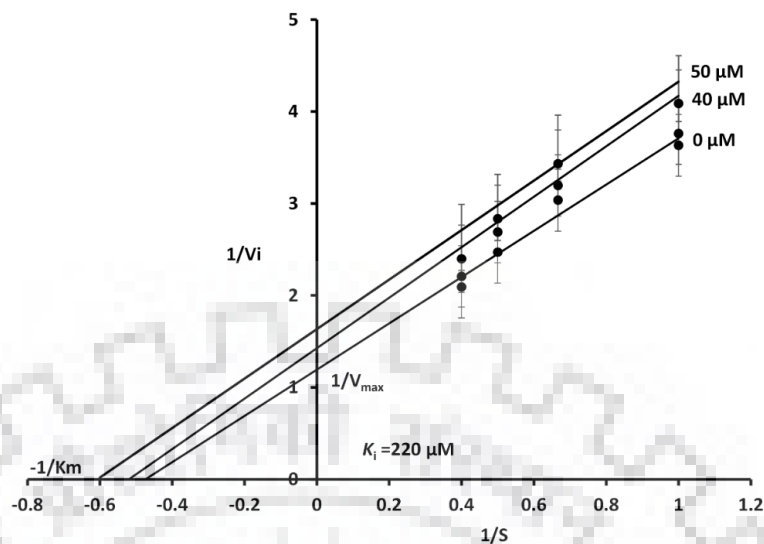


Figure 4.4.3.2: Determination of Inhibition constant for TPCK: The lineweaver-burk plot of $1/V_i$ against $1/[S]$ (double reciprocal of V_i and S) was generated. The slope of the plot gives K_m/V_{max} , x-intercept gives $1/V_{max}$ and y-intercept gives $-1/K_m$. As seen from the plot, with the increase in the inhibitor concentration the value of apparent K_m decreases (y-intercept moves left) V_{max} decreases. The TPCK showed uncompetitive mode of inhibition for CVCP protease activity. The value of inhibition constant (K_i) obtained for TPCK was $220 \mu\text{M}$.

4.5 Conclusion

Chikungunya fever is an arthropod-borne, re-emerging viral disease responsible for major global health concern. The evolution of new mosquito vector (*Aedes albopictus*) for CHIKV due to mutation of Ala226 to Val in E1 glycoprotein, shift of disease presentation from arthralgia to encephalitis and consequently re-emergence of CHIKV as an epidemic; demands proper and immediate attention. Presently, there is no specific medicine or vaccine available against CHIKV infection. Although, investigations on numerous antiviral compounds, immunomodulatory drugs and monoclonal antibodies that could be used for treatment against CHIKV are under process. Therefore, there is an immediate need of development of novel drugs and efficient therapeutics against CHIKV infection.

Alphavirus CP is a multi-functional protein comprising of two domains, RNA binding N-terminal domain and autoproteolytic C-terminal domain. The N-terminal domain is highly disoriented with high degree positive charge and is not well conserved. The N-terminal domain is responsible for binding to RNA genome, dimerization of CP, nuclear trafficking and downregulation of host transcription. The C-terminal domain of alphavirus CP is highly conserved and is a chymotrypsin-like serine protease. It has *cis*-proteolytic activity that cuts at

W/S scissile bond and separates itself from nascent structural polyprotein. The C-terminal protease domain of CP plays a critical role in the processing of structural polyprotein, cytoplasmic trafficking and the viral life cycle. After the proteolytic cleavage, the C-terminal tryptophan residue of CP binds to S1 specificity pocket and blocks the further protease activity. Recent studies have revealed that the basic molecular architecture of the active site and residues involved in catalysis are highly conserved among the serine proteases including the CP of alphaviruses (His139, Asp161 and Ser213 CVCP residues) (Aggarwal et al., 2012). In 2008, Morillas et al., reported the esterase activity of SFCP truncated protein. But till date, *in vitro trans*-protease activity of alphavirus CP has not reported and consequently CP protease activity has not been targeted for development of drugs against members of alphavirus genus.

In the present study the last two residues (₂₆₀E and W₂₆₁) from the C-terminal CP was deleted and the *trans*-protease activity was evaluated by using FRET based fluorogenic peptide substrate. The purified CVCP truncated protein was found to be fully active. The sequence of substrate peptide was taken from CVCP protease cleavage site encompassing W/S scissile bond. The CVCP protease cleaves the fluorogenic substrate into two fragments separating DABCYL and EDANS. This results in reduction of FRET signal and increase in the fluorescence emission intensity for the donor fluorophore. The upsurge in the reaction velocity has been observed with increase in the substrate concentration until it gets saturated. Kinetic parameters using fluorogenic peptide substrates for the chikungunya virus capsid protease were estimated. The K_m and K_{cat}/K_m values were determined and were found to be $0.754 \pm 0.11 \mu\text{M}$ and $1.72 \times 10^5 \text{ M}^{-1} \text{ min}^{-1}$, respectively. The effect of serine protease inhibitors TPCK and BHC on CVCP protease activity has also been determined. The value of dissociation constant (K_i) obtained for BHC and TPCK were $115 \mu\text{M}$ and $220 \mu\text{M}$, respectively. The value of Z' factor calculated for the FRET assay was 0.64 that further validates the assay for carrying out screening of compound libraries/FDA approved drugs against CVCP protease. In conclusion, this FRET based *in vitro* protease assay can be used as HTS method for testing and screening inhibitors against the proteolytic activity of CHIKV capsid protease. Since, the elementary molecular architecture of the active site and the catalytic triad is highly conserved among the members of serine protease family including alphaviruses CPs, this assay can be used for screening of anti-alphaviral inhibitors.

References

- Abdelnabi, R., Neyts, J., Delang, L. (2017). Chikungunya virus infections: time to act, time to treat. *Curr Opin Virol.* 24, 25-30.
- Acheson, N. H. & Tamm, I. (1967). Replication of Semliki Forest virus: an electron microscopic study. *Virology* 32, 128–143.
- Adams, P.D., Afonine, P.V., BunkAczi, G.b., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* 66, 213-221.
- Agarwal, A., Singh, A.K., Sharma, S., Soni, M., Thakur, A.K., Gopalan, N., Parida, M.M., Rao, P.V.L., Dash, P.K. (2013). Application of Real-time RT-PCR in vector surveillance and assessment of replication kinetics of an emerging novel ECSA genotype of Chikungunya virus in *Aedes aegypti*. *J. Virol. Methods* 193, 419-425.
- Aggarwal M, Dhindwal S, Kumar P, Kuhn RJ, Tomar S. (2014). trans-Protease activity and structural insights into the active form of the alphavirus capsid protease. *J. Virol.* 88, 12242-12253.
- Aggarwal M, Tapas S, Siwach A, Kumar P, Kuhn RJ, Tomar S. (2012). Crystal structure of aura virus capsid protease and its complex with dioxane: new insights into capsid-glycoprotein molecular contacts. *PloS one* 7:e51288
- Aggarwal, M., Ramanjit, K., Saha, A., Mudgal, R., Yadav, R., Dash, P.K., Parida, M., Kumar, P., Tomar, S. (2017). Evaluation of antiviral activity of piperazine against Chikungunya virus targeting hydrophobic pocket of alphavirus capsid protein. *Antiviral Res.* (*Communicated*).
- Aggarwal, M., Sharma, R., Kumar, P., Parida, M., Tomar, S. (2015). Kinetic characterization of trans-proteolytic activity of Chikungunya virus capsid protease and development of a FRET-based HTS assay. *Scientific reports* 5.
- Ahola, T. & Kääriäinen, L. (1995). Reaction in alphavirus mRNA capping: formation of a covalent complex of nonstructural protein nsP1 with 7-methyl-GMP. *Proc. Natl. Acad. Sci. USA* 92, 507–511.
- Aliperti, G. & Schlesinger, M. J. (1978). Evidence for an autoprotease activity of Sindbis virus capsid protein. *Virology* 90, 366–369.

Allaire, M., Chernaia, M. M., Malcolm, B. A. & James, M. N. G. (1994). Picornaviral 3C cysteine proteinases have a fold similar to the chymotrypsin-like serine proteinases. *Nature* 369, 72–76.

Anand, R., Maksimoska, J., Pagano, N., Wong, E.Y., Gimotty, P.A., Diamond, S.L., Meggers, E. and Marmorstein, R. (2009). Toward the development of a potent and selective organoruthenium mammalian sterile 20 kinase inhibitor. *Journal of medicinal chemistry*, 52(6), pp.1602-1611.

Angliker, H., Neumann, U., Molloy, S. S., Thomas, G. (1995). Internally quenched fluorogenic substrate for furin. *Anal. Biochem.* 224, 409–412.

Balasubramanian, A., Manzano, M., Teramoto, T., Pilankatta, R., & Padmanabhan, R. (2016). High-throughput screening for the identification of small-molecule inhibitors of the flaviviral protease. *Antiviral research*, 134, 6-16.

Baron, S., Schmaljohn, A.L., McClain, D. (1996). Alphaviruses (togaviridae) and flaviviruses (flaviviridae).

Barth, B. U, Suomalainen, M., Liljeström, P. & Garoff, H. (1992). Alphavirus assembly and entry: role of the cytoplasmic tail of the E1 spike subunit. *J. Virol.* 66, 7560–7564.

Blanchard, J. E., Elowe, N. H., Huitema, C., Fortin, P. D., Cechetto, J. D., Eltis, L. D., & Brown, E. D. (2004). High-throughput screening identifies inhibitors of the SARS coronavirus main proteinase. *Chemistry & biology* 11, 1445-1453.

BODE, W., TURK, D., & STURZEBECKER, J. (1990). Geometry of binding of the benzamide- and arginine-based inhibitors N α -(2-naphthyl-sulphonyl-glycyl)-dl-p-amidinophenylalanyl-piperidine (NAPAP) and (2R, 4R)-4-methyl-1-[N α -(3-methyl-1, 2, 3, 4-tetrahydro-8-quinolinesulphonyl)-l-arginyl]-2-piperidine carboxylic acid (MQPA) to human α -thrombin. *The FEBS Journal* 193, 175-182.

Brown, D.T. (1980). Assembly of alphaviruses. Academic Press, Inc., New York.

Carter, P. & Wells, J. A. (1988). Dissection of the catalytic triad of a serine protease. *Nature* 332, 564–568.

Cavrini, F., Gaibani, P., Pierro, A.M., Rossini, G., Landini, M.P., Sambri, V. (2009). Chikungunya: an emerging and spreading arthropod-borne viral disease. *The Journal of Infection in Developing Countries* 3, 744-752.

Chandak, N.H., Kashyap, R.S., Kabra, D., Karandikar, P., Saha, S.S., Morey, S.H., Purohit, H.J., Taori, G.M., Dagainawala, H.F. (2009). Neurological complications of Chikungunya virus infection. *Neurol. India* 57, 177.

Cheng, R. H., Kuhn, R. J., Olson, N. H., Rossmann, M. G., Choi, H. K., Smith, T. J. & Baker, T. S. (1995). Nucleocapsid and glycoprotein organization in an enveloped virus. *Cell* 80, 621–630.

Choi, H. K., Lee, S., Zhang, Y. P., McKinney, B. R., Wengler, G., Rossmann, M. G. & Kuhn, R. J. (1996). Structural analysis of Sindbis virus capsid mutants involving assembly and catalysis. *J. Mol. Biol.* 262, 151–167.

Choi, H. K., Lu, G., Lee, S., Wengler, G. & Rossmann M. G. (1997). Structure of Semliki Forest Virus Core Protein. *Proteins: Struct. Funct. Genet.* 27, 345–359.

Choi, H. K., Tong, L., Minor, W., Dumas, P., Boege, U., Rossmann, M. G. & Wengler, G. (1991) Structure of Sindbis virus core protein reveals a chymotrypsin-like serine proteinase and the organization of the virion. *Nature* 354, 37–43.

Choi, K. H. (2012). Viral polymerases. In *Viral Molecular Machines* (pp. 267-304). Springer US.

Choi, K. H., & Rossmann, M. G. (2009). RNA-dependent RNA polymerases from Flaviviridae. *Current opinion in structural biology*, 19(6), 746-751.

Chusri, S., Siripaitoon, P., Hirunpat, S., Silpapojakul, K. (2011). Case reports of neuro-Chikungunya in southern Thailand. *The American journal of tropical medicine and hygiene* 85, 386-389.

Corrêa, P.R.C., da Silveira Pinto, A., de Fátima Silva, G.D., Vieira Filho, S.A., Duarte, L.P., Yamada-Ogatta, S.F. (2011). In vitro activity of 2-pyridinecarboxylic acid against trypanosomes of the subgenus *Schizotrypanum* isolated from the bat *Phyllostomus hastatus*-doi: 10.4025/actasciobiolsci. v33i4. 6482. *Acta Scientiarum. Biological Sciences* 33, 437-443.

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall Iii, W.B., Snoeyink, J., Richardson, J.S. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 35, 375-383.

DeLano, W.L., (2002). The PyMOL molecular graphics system. <http://pymol.org>.

Dhindwal, S., Kesari, P., Singh, H., Kumar, P., Tomar, S. (2016). Conformer and pharmacophore based identification of peptidomimetic inhibitors of chikungunya virus nsP2 protease. *J. Biomol. Struct. Dyn.*, 1-18.

Emini, E.A., Hughes, J.V., Perlow, D., Boger, J. (1985). Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J. Virol.* 55, 836-839.

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

Feng, Y. and LeBlanc, M.H. (2003). Treatment of hypoxic-ischemic brain injury in newborn rats with TPCK 3 h after hypoxia decreases caspase-9 activation and improves neuropathologic outcome. *Developmental neuroscience*, 25, pp.34-40.

Fernandez-Pol, J.A., Fernandez-Pol, S. (2008). Method to control dengue viruses in humans by picolinic acid and derivatives thereof. Google Patents.

Fernandez-Pol, J.A., Klos, D.J., Hamilton, P.D. (2000). Antiviral, cytotoxic and apoptotic activities of picolinic acid on human immunodeficiency virus-1 and human herpes simplex virus-2 infected cells. *Anticancer Res.* 21, 3773-3776.

Forsell, K., Griffiths, G. & Garoff, H. (1996). Preformed cytoplasmic nucleocapsids are not necessary for alphavirus budding. *EMBO J.* 15, 6495-6505.

Forsell, K., Suomalainen, M. & Garoff, H. (1995). Structure-function relation of the NH₂-terminal domain of the Semliki Forest virus capsid protein. *J. Virol.* 69, 1556-1563.

Forsell, K., Xing, L., Kozlovska, T., Cheng, R. H. & Garoff, H. (2000). Membrane proteins organize a symmetrical virus. *EMBO J.* 19, 5081-5091.

Frolov, I., & Schlesinger, S. (1994). Translation of Sindbis virus mRNA: Effects of sequences downstream of the initiating codon. *J. Virol.* 68, 8111-8117.

Frolov, I., Hardy, R. & Rice, C. M. (2001). Cis-acting RNA elements at the 5' end of Sindbis virus genome RNA regulate minus- and plus-strand RNA synthesis. *RNA* 7, 1638-1651.

Fuller, S. D. (1987). The T=4 envelope of Sindbis virus is organized by interactions with a complementary T=3 capsid. *Cell* 48, 923-934.

Gaedigk-Nitschko, K. & Schlesinger, M. J. (1990). The Sindbis virus 6K protein can be detected in virions and is acylated with fatty acids. *Virology* 175, 274-281.

Gaedigk-Nitschko, K. & Schlesinger, M.J. (1991). Site-directed mutations in Sindbis virus E2 glycoprotein's cytoplasmic domain and the 6K protein lead to similar defects in virus assembly and budding. *Virology* 183, 206–214.

Gaedigk-Nitschko, K., Ding, M. X., Levy, M. A. & Schlesinger, M. J. (1990) Site-directed mutations in the Sindbis virus 6K protein reveal sites for fatty acylation and the underacylated protein affects virus release and virion structure. *Virology* 175, 282–291.

Garmashova, N., Gorchakov, R., Volkova, E., Paessler, S., Frolova, E., Frolov, I. (2007). The Old World and New World alphaviruses use different virus-specific proteins for induction of transcriptional shutoff. *J. Virol.* 81, 2472-2484.

Garoff, H., & Simons, K. (1974). Location of the spike glycoproteins in the Semliki Forest virus membrane. *Proc. Natl. Acad. Sci. USA* 71, 3988–3992.

Garoff, H., Frischauf, A. M., Simons, K., Lehrach, H., & Delius, H. (1980). The capsid protein of Semliki Forest virus has clusters of basic amino acids and prolines in its amino-terminal region. *Proc. Natl. Acad. Sci. USA* 77, 6376–6380.

Garoff, H., Simons, K., & Dobberstein, B. (1978). Assembly of the Semliki Forest virus membrane glycoproteins in the membrane of the endoplasmic reticulum in vitro. *J. Mol. Biol.* 124, 587–600.

Garoff, H., Sjoberg, M., & Cheng, R. H. (2004). Budding of alphaviruses. *Virus research* 106, 103-116.

Garoff, H., Wilschut, J., Liljestrom, P., Wahlberg, J. M., Bron, R., Suomalainen, M., Smyth, J., Salminen, A. Barth, B. U., Zhao, H., Forsell, K. & Ekström, M. (1994). Assembly and entry mechanisms of Semliki Forest virus. *Arch. Virol. Suppl.* 9, 329–338.

Geigenmüller-Gnirke, U., Nitschko, H. & Schlesinger, S. (1993). Deletion analysis of the capsid protein of Sindbis virus: identification of the RNA binding region. *J. Virol.* 67, 1620–1626.

Gerardin, P., Guernier, V., Perrau, J.L., Fianu, A., Le Roux, K., Grivard, P., Michault, A., De Lamballerie, X., Flahault, A., Favier, F.o. (2008). Estimating Chikungunya prevalence in La Reunion Island outbreak by serosurveys: two methods for two critical times of the epidemic. *BMC Infect. Dis.* 8, 99.

Goh, L.Y.H., Hobson-Peters, J., Prow, N.A., Baker, K., Piyasena, T.B.H., Taylor, C.T., Rana, A., Hastie, M.L., Gorman, J.J., Hall, R.A. (2015). The Chikungunya virus capsid protein contains

linear B cell epitopes in the N-and C-terminal regions that are dependent on an intact C-terminus for antibody recognition. *Viruses* 7, 2943-2964.

Gómez de Cedrón, M., Ehsani, N., Mikkola, M. L., García, J. A. & Kääriäinen, L. (1999). RNA helicase activity of Semliki Forest virus replicase protein NSP2. *FEBS Lett.* 448, 19–22.

Gonzalez, M. E. & Carrasco, L. (2003). Viroporins. *FEBS Lett.* 552, 28–34.

Hahn, C. S. & Strauss, J. H. (1990). Site-directed mutagenesis of the proposed catalytic amino acids of the Sindbis virus capsid protein autoprotease. *J. Virol.* 64, 3069–3073.

Hahn, C. S., Strauss, E. G. & Strauss, J. H. (1985). Sequence analysis of three Sindbis virus mutants temperature-sensitive in the capsid autoprotease. *Proc. Natl. Acad. Sci. USA* 82, 4648–4652.

Hahn, C.S., Lustig, S., Strauss, E.G., Strauss, J.H. (1988). Western equine encephalitis virus is a recombinant virus. *Proceedings of the National Academy of Sciences* 85, 5997-6001.

Hahn, Y. S., Grakoui, A., Rice, C. M., Strauss, E. G. & Strauss, J. (1989). Mapping of RNA-temperature-sensitive mutants of Sindbis virus: complementation group F mutants have lesions in nsP4. *J. Virol.* 63, 1194–1202.

Hardy, R. W. (2006). The role of the 3' terminus of the Sindbis virus genome in minus-strand initiation site selection. *Virology* 345, 520–531.

Hardy, W. R. & Strauss, J. H. (1988). Processing the nonstructural polyproteins of Sindbis virus: study of the kinetics in vivo by using monospecific antibodies. *J. Virol.* 62, 998–1007.

Hardy, W. R. & Strauss, J. H. (1989). Processing the nonstructural polyproteins of sindbis virus: nonstructural proteinase is in the C-terminal half of nsP2 and functions both in cis and in trans. *J. Virol.* 63, 4653–64.

Hedstrom, L. (2002). Serine protease mechanism and specificity. *Chemical reviews*, 102, 4501-4524.

Helenius, A., Kartenbeck, J., Simons, K. & Fries, E. (1980). On the entry of Semliki Forest virus into BHK-21 cells. *J. Cell Biol.* 84, 404–420.

Hong, E. M., Perera, R. & Kuhn, R. J. (2006). Alphavirus Capsid Protein Helix I Controls a Checkpoint in Nucleocapsid Core Assembly. *J. Virol.* 80, 8848–8855.

Ivanova, L. & Schlesinger, M. J. (1993). Site-directed mutations in the Sindbis virus E2 glycoprotein identify palmitoylation sites and affect virus budding. *J. Virol.* 67, 2546–2551.

- Jadav, S.S., Sinha, B.N., Hilgenfeld, R., Pastorino, B., de Lamballerie, X., Jayaprakash, V. (2015). Thiazolidone derivatives as inhibitors of chikungunya virus. *European journal of medicinal chemistry* 89, 172-178.
- Jain, D., & Lamour, V. (2010). Computational tools in protein crystallography. *Computational Biology*, 129-156.
- Jain, D. (2016). Cloning, expression, purification, crystallization and initial crystallographic analysis of FlcN from *Pseudomonas aeruginosa*. *Acta Crystallographica Section F: Structural Biology Communications*, 72(2), 135-138.
- Jain, S. K., DeCandido, S. & Kielian, M. (1991). Processing of the p62 envelope precursor protein of Semliki Forest virus. *J. Biol. Chem.* 266, 5756–5761.
- Jose, J., Przybyla, L., Edwards, T. J., Perera, R., Burgner, J. W. & Kuhn, R. J. (2012). Interactions of the cytoplasmic domain of Sindbis virus E2 with nucleocapsid cores promote alphavirus budding. *J. Virol.* 86, 2585–2599.
- Jose, J., Snyder, J. E., & Kuhn, R. J. (2009). A structural and functional perspective of alphavirus replication and assembly. *Future microbiology*, 4, 837-856.
- Justman, J., Klimjack, M. R. & Kielian, M. (1993). Role of spike protein conformational changes in fusion of Semliki Forest virus. *J. Virol.* 67, 7597–7607.
- Kail, M., Hollinshead, M., Ansorge, W., Pepperkok, R., Frank, R., Griffiths, G. & Vaux, D. (1991). The cytoplasmic domain of alphavirus E2 glycoprotein contains a short linear recognition signal required for viral budding. *EMBO J.* 10, 2343–2351.
- Kam, Y.-W., Lee, W.W.L., Simarmata, D., Le Grand, R., Tolou, H., Merits, A., Roques, P., Ng, L.F.P. (2014). Unique epitopes recognized by antibodies induced in Chikungunya virus-infected non-human primates: implications for the study of immunopathology and vaccine development. *PLoS One* 9, e95647.
- Kamer G, Argos P. (1984). Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. *Nucleic Acids Res.*; 12, 7269–7282.
- Karplus, P.A., Schulz, G.E. (1985). Prediction of chain flexibility in proteins. *Naturwissenschaften* 72, 212-213.
- Kielian, M. & Helenius, A. (1985). pH-induced alterations in the fusogenic spike protein of Semliki Forest virus. *J. Cell Biol.* 101, 2284–2291.

- Kielian, M. (1995). Membrane fusion and the alphavirus life cycle. *Adv. Virus Res.* 45, 113–151.
- Kielian, M. (2002). Structural surprises from the flaviviruses and alphaviruses. *Mol. Cell* 9, 454–456.
- Kim, H.Y., Kuhn, R.J., Patkar, C., Warriar, R., Cushman, M. (2007). Synthesis of dioxane-based antiviral agents and evaluation of their biological activities as inhibitors of Sindbis virus replication. *Bioorg. Med. Chem.* 15, 2667-2679.
- Kim, H.Y., Patkar, C., Warriar, R., Kuhn, R., Cushman, M. (2005). Design, synthesis, and evaluation of dioxane-based antiviral agents targeted against the Sindbis virus capsid protein. *Bioorg. Med. Chem. Lett.* 15, 3207-3211.
- Klimstra, W. B., Nangle, E. M., Smith, M. S., Yurochko, A. D. & Ryman, K. D. (2003). DC-SIGN and L-SIGN can act as attachment receptors for alphaviruses and distinguish between mosquito cell- and mammalian cell-derived viruses. *J. Virol.* 77, 12022–12032.
- Klimstra, W. B., Ryman, K. D. & Johnston, R. E. (1998). Adaptation of sindbis virus to BHK cells selects for use of heparan sulfate as an attachment receptor. *J. Virol.* 72, 7357–7366.
- Knight, C. G., Willenbrock, F. & Murphy, G. (1992). A novel coumarin-labelled peptide for sensitive continuous assays of the matrix metalloproteinases. *FEBS Lett.* 296, 263–266.
- Knight, R. L., Schultz, K. L., Kent, R. J., Venkatesan, M. & Griffin, D. E. (2009). Role of N-linked glycosylation for Sindbis virus infection and replication in vertebrate and invertebrate systems. *J. Virol.* 83, 5640–5647.
- Kolaskar, A.S., Tongaonkar, P.C. (1990). A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett.* 276, 172-174.
- Kondor-Koch, C., Burke, B. & Garoff, H. (1983). Expression of Semliki Forest virus proteins from cloned complementary DNA. I. The fusion activity of the spike glycoprotein. *J. Cell Biol.* 97, 644–651.
- Kostyuchenko, V.A., Jakana, J., Liu, X., Haddow, A.D., Aung, M., Weaver, S.C., Chiu, W., Lok, S.-M. (2011). The structure of barmah forest virus as revealed by cryo-electron microscopy at a 6-angstrom resolution has detailed transmembrane protein architecture and interactions. *J. Virol.* 85, 9327-9333.

- Kuhn, R. J., Hong, Z. & Strauss, J. H. (1990). Mutagenesis of the 3' nontranslated region of Sindbis virus RNA. *J. Virol.* 64, 1465–1476.
- Kumar, C.V.M.N., Gopal, D.V.R.S. (2010). Reemergence of chikungunya virus in Indian Subcontinent. *Indian J. Virol.* 21, 8-17.
- Kumar, A., Sharma, J., Mohanty, A. K., Grover, S. & Batish, V. K. (2006). Purification and characterization of milk clotting enzyme from goat (*Capra hircus*). *Comp. Biochem. Physiol. Part B: Biochem. Mol. Biol.* 145, 108–113.
- Larrieu, S., Pouderoux, N., Pistone, T., Filleul, L., Receveur, M. C., Sissoko, D., Ezzedine, K. & Malvy, D. (2010). Factors associated with persistence of arthralgia among Chikungunya virus-infected travellers: report of 42 French cases. *J. Clin. Virol.* 47, 85–88.
- Larsen, J.E.P., Lund, O., Nielsen, M. (2006). Improved method for predicting linear B-cell epitopes. *Immunome research* 2, 2.
- Lee, S., Owen, K.E., Choi, H.K., Lee, H., Lu, G., Wengler, G., Brown, D.T., Rossmann, M.G. and Kuhn, R.J. (1996). Identification of a protein binding site on the surface of the alphavirus nucleocapsid and its implication in virus assembly. *Structure* 4, 531–41
- Lee, H. & Brown, D. T. (1994). Mutations in an exposed domain of Sindbis virus capsid protein result in the production of noninfectious virions and morphological variants. *Virology* 202, 390–400.
- Lee, S., Owen, K. E., Choi, H. K., Lee, H., Lu, G., Wengler, G., Brown, D. T., Rossmann, M. G. & Kuhn, R. J. (1996). Identification of a protein binding site on the surface of the alphavirus nucleocapsid and its implications in virus assembly. *Structure* 4, 531–541.
- Lemm JA, Rumenapf T, Strauss EG, Strauss JH, Rice CM. (1994). Polypeptide requirements for assembly of functional Sindbis virus replication complexes: a model for the temporal regulation of minus- and plus-strand RNA synthesis. *EMBO J*, 13, 2925–2934.
- Levine, B., Jiang, H. H., Kleeman, L. & Yang, G. (1996). Effect of E2 envelope glycoprotein cytoplasmic domain mutations on Sindbis virus pathogenesis. *J. Virol.* 70, 1255–1260.
- Li, G. & C. M. Rice. (1993). The signal for translational readthrough of a UGA codon in Sindbis virus RNA involves a single cytidine residue immediately downstream of the termination codon. *J. Virol.* 67, 5062–5067.

- Li, G. & Rice, C. M. (1989). Mutagenesis of the in-frame opal termination codon preceding nsP4 of Sindbis virus: Studies of translational readthrough and its effect on virus replication. *J. Virol.* 63, 1326–1337.
- Li, L., Jose, J., Xiang, Y., Kuhn, R. J. & Rossmann, M. G. (2010). Structural changes of envelope proteins during alphavirus fusion. *Nature* 468, 705–708.
- Ligon, B. L. (2006). Reemergence of an unusual disease: the chikungunya epidemic. *Semin. Pediatr. Infect. Dis.* 17, 99–104.
- Liljeström, P. & Garoff, H. (1991). Internally located cleavable signal sequences direct the formation of Semliki Forest virus membrane proteins from a polyprotein precursor. *J. Virol.* 65, 147–154.
- Liljestrom, P., Lusa, S., Huylebroeck, D. & Garoff, H. (1991). In vitro mutagenesis of a full-length cDNA clone of Semliki Forest virus: the small 6,000-molecular-weight membrane protein modulates virus release. *J. Virol.* 65, 4107–4113.
- Linger BR, Kunovska L, Kuhn RJ, Golden BL. (2004). Sindbis virus nucleocapsid assembly: RNA folding promotes capsid protein dimerization. *RNA.* 10, 128–138
- Liu, L. N., Lee, H., Hernandez, R. & Brown, D. T. (1996). Mutations in the endo domain of Sindbis virus glycoprotein E2 block phosphorylation, reorientation of the endo domain, and nucleocapsid binding. *Virology* 222, 236–246.
- Liu, N. & Brown, D. T. (1993). Phosphorylation and dephosphorylation events play critical roles in Sindbis virus maturation. *Virology* 196, 703–711.
- Liu, N. & Brown, D. T. (1993). Transient translocation of the cytoplasmic (endo) domain of a type I membrane glycoprotein into cellular membranes. *J. Cell Biol.* 120, 877–883.
- Lobigs, M., Zhao, H. X. & Garoff, H. (1990). Function of Semliki Forest virus E3 peptide in virus assembly: replacement of E3 with an artificial signal peptide abolishes spike heterodimerization and surface expression of E1. *J. Virol.* 64, 4346–4355.
- Loewy, A., Smyth, J., von Bonsdorff, C. H., Liljestrom, P. & Schlesinger, M. J. (1995). The 6-kilodalton membrane protein of Semliki Forest virus is involved in the budding process. *J. Virol.* 69, 469–475.
- Lopez, S., Yao, J. S., Kuhn, R. J., Strauss, E. G. & Strauss, J. H. (1994). Nucleocapsid-glycoprotein interactions required for assembly of alphaviruses. *J. Virol.* 68, 1316–1323.

- Lucas-Hourani, M., Lupan, A., Despres, P., Thoret, S., Pamard, O., et al. (2013). A phenotypic assay to identify Chikungunya virus inhibitors targeting the nonstructural protein nsP2. *J. Biomol. Screen* 18, 172–179.
- Lulla, V., Frolova, E.I., Frolov, I. (2013). The amino-terminal domain of alphavirus capsid protein is dispensable for viral particle assembly but regulates RNA encapsidation through cooperative functions of its subdomains. *J. Virol.* 87, 12003-12019.
- Lusa, S., Garoff, H. & Liljestrom, P. (1991). Fate of the 6K membrane protein of Semliki Forest virus during virus assembly. *Virology* 185, 843–846.
- Luthy, R., Bowie, J. U. & Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. *Nature* 356, 83–85.
- Madan, V., Castello, A. & Carrasco, L. (2008). Viroporins from RNA viruses induce caspase-dependent apoptosis. *Cell. Microbiol.* 10, 437–451.
- Madan, V., Sanz, M. A. & Carrasco, L. (2005). Requirement of the vesicular system for membrane permeabilization by Sindbis virus. *Virology* 332, 307–315.
- Malet, H., Gould, E. A., Jamal, S., et al. (2009). The crystal structures of Chikungunya and Venezuelan equine encephalitis virus nsP3 macro domains define a conserved adenosine binding pocket. *J. Virol.* 83, 6534–6545.
- Mancini, E. J., Clarke, M., Gowen, B. E., Rutten, T. & Fuller, S. D. (2000). Cryo-electron microscopy reveals the functional organization of an enveloped virus, Semliki Forest virus. *Mol. Cell* 5, 255–266.
- Matthews, D. A., Smith, W. W., Ferre, R. A., Condon, B., Budahazi, G., Sisson, W., Villafranca, J. E., Janson, C. A., McElroy, H. E., Gribskor, C. L. & Worland, S. (1994). Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein. *Cell* 77, 761–771.
- Mavalankar, D., Shastri, P., Bandyopadhyay, T., Parmar, J., Ramani, K.V. (2008). Increased mortality rate associated with chikungunya epidemic, Ahmedabad, India. *Emerg. Infect. Dis.* 14, 412-415.
- Melancon, P. & Garoff, H. (1987). Processing of the Semliki Forest virus structural polyprotein: role of the capsid protease. *J. Virol.* 61, 1301–1309.

Mellman, I., Fuchs, R. & Helenius, A. (1986). Acidification of the endocytic and exocytic pathways. *Annu. Rev. Biochem.* 55, 663–700.

Melton, J. V., Ewart, G. D., Weir, R. C., Board, P. G., Lee, E. & Gage, P. W. (2002). Alphavirus 6K proteins form ion channels. *J. Biol. Chem.* 277, 46923–46931.

Metallo, S. J. (2010). Intrinsically disordered proteins are potential drug targets. *Curr. Opin. Chem. Biol.* 14, 481–488

Metsikkö, K. & Garoff, H. (1990). Oligomers of the cytoplasmic domain of the p62/E2 membrane protein of Semliki Forest virus bind to the nucleocapsid in vitro. *J. Virol.* 64, 4678–4683.

Mi, S. & Stollar, V. (1991). Expression of Sindbis virus nsP1 and methyltransferase activity in *Escherichia coli*. *Virology* 184, 423–427.

Mittoo, S., Sundstrom, L. E. & Bradley, M. (2003). Synthesis and evaluation of fluorescent probes for the detection of calpain activity. *Anal. Biochem.* 319, 234–238.

Mohanty, A. K., Mukhopadhyay, U. K., Kaushik, J. K., Grover, S. & Batish, V. K. (2003). Isolation, purification and characterization of chymosin from riverine buffalo (*Bubalus bubalis*). *J. dairy res.* 70, 37–43.

Morillas, M., Eberl, H., Allain, F. H. T., Glockshuber, R. & Kuennemann, E. (2008). Novel Enzymatic Activity Derived from the Semliki Forest Virus Capsid Protein. *J. Mol. Biol.* 376, 721–735

Mueller, N. H., Pattabiraman, N., Ansarah-Sobrinho, C., Viswanathan, P., Pierson, T. C. & Padmanabhan, R. (2008). Identification and biochemical characterization of small molecule inhibitors of West Nile Virus serine protease by a high throughput screen. *Antimicrob. Agents Chemother.* 52, 3385–3393.

Morrison, T.E. (2014). Reemergence of chikungunya virus. *J. Virol.* 88, 11644–11647.

Mukhopadhyay, S., Chipman, P. R., Hong, E. M., Kuhn, R. J. & Rossmann, M. G. (2002). In vitro-assembled alphavirus core-like particles maintain a structure similar to that of nucleocapsid cores in mature virus. *J. Virol.* 76, 11128–11132.

Mukhopadhyay, S., Zhang, W., Gabler, S., Chipman, P.R., Strauss, E.G., Strauss, J.H., Baker, T.S., Kuhn, R.J., Rossmann, M.G. (2006). Mapping the structure and function of the E1 and E2 glycoproteins in alphaviruses. *Structure* 14, 63-73.

Myles, K. M., Pierro, D. J. & Olson, K. E. (2003). Deletions in the putative cell receptor-binding domain of Sindbis virus strain MRE16 E2 glycoprotein reduce midgut infectivity in *Aedes aegypti*. *J. Virol.* 77, 8872–8881.

Nimmannitya, S., Halstead, S. B., Cohen, S. N. & Margiotta, M. R. (1969). Dengue and chikungunya virus infection in man in Thailand, 1962–1964. I. Observations on hospitalized patients with hemorrhagic fever. *Am. J. Trop. Med. Hyg.* 18, 954–971.

Odegard, A. L., Kwan, M. H., Walukiewicz, H. E., Banerjee, M., Schneemann, A. & Johnson, J. E. (2009). Low endocytic pH and capsid protein autocleavage are critical components of Flock House virus cell entry. *J. Virol.* 83, 8628–8637.

Otto, H. H.; Schirmeister, T. (1997). *Chem. Rev.*, 97, 133

Otwinowski, Z., Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. 307-326.

Ou, J. H., Strauss, E. G. & Strauss, J. H. (1983). The 5'-terminal sequences of the genomic RNAs of several alphaviruses. *J. Mol. Biol.* 168, 1–15.

Owen, K. E. & Kuhn, R. J. (1996). Identification of a region in the Sindbis virus nucleocapsid protein that is involved in specificity of RNA encapsidation. *J. Virol.* 70, 2757–2763.

Owen, K. E. & Kuhn, R. J. (1997). Alphavirus budding is dependent on the interaction between the nucleocapsid and hydrophobic amino acids on the cytoplasmic domain of the E2 envelope glycoprotein. *Virology* 230, 187–196.

Padbidri, V.S., Gnaneswar, T.T. (1978). Epidemiological investigations of chikungunya epidemic at Barsi, Maharashtra state, India. *Journal of hygiene, epidemiology, microbiology, and immunology* 23, 445-451.

Paredes, A. M., Brown, D. T., Rothnagel, R., Chiu, W., Schoepp, R. J., Johnston, R. E. & Prasad, B. V. (1993). Three-dimensional structure of a membrane-containing virus. *Proc. Natl Acad. Sci. USA* 90, 9095–9099.

Paredes, A. M., Heidner, H., Thuman-Commike, P., Prasad, B. V., Johnston, R. E. & Chiu, W. (1998). Structural localization of the E3 glycoprotein in attenuated Sindbis virus mutants. *J. Virol.* 72, 1534–1541.

Paredes, A. M., Simon, M. N. & Brown, D. T. (1992). The mass of the Sindbis virus nucleocapsid suggests it has T = 4 icosahedral symmetry. *Virology* 187, 329–332.

- Park, E. & Griffin, D. E. (2009). Interaction of Sindbis virus nonstructural protein 3 with poly (ADP-ribose) polymerase-1 in neuronal cells. *J. Gen. Virol.* 90, 2073–2080.
- Park, E. & Griffin, D. E. (2009). The nsP3 macro domain is important for Sindbis virus replication in neurons and neurovirulence in mice. *Virology* 322, 305–314.
- Parker, J.M.R., Guo, D., Hodges, R.S. (1986). New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry* 25, 5425-5432.
- Peränen, J., Laakkonen, P., Hyvönen, M. & Kääriäinen, L. (1995). The alphavirus replicase protein nsP1 is membrane-associated and has affinity to endocytic organelles. *Virology* 208, 610–620.
- Perera, R., Navaratnarajah, C. & Kuhn, R. J. (2003). A heterologous coiled coil can substitute for helix I of the Sindbis virus capsid protein. *J. Virol.* 77, 8345–8353.
- Perera, R., Owen, K. E., Tellinghuisen, T. L., Gorbalenya, A. E. & Kuhn, R. J. (2001). Alphavirus nucleocapsid protein contains a putative coiled coil alpha-helix important for core assembly. *J. Virol.* 75, 1–10.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry* 25, 1605-1612.
- Pletnev, S. V., Zhang, W., Mukhopadhyay, S., Fisher, B. R., Hernandez, R., Brown, D. T., Baker, T. S., Rossmann, M. G. & Kuhn, R. J. (2001). Locations of carbohydrate sites on alphavirus glycoproteins show that E1 forms an icosahedral scaffold. *Cell* 105, 127–136.
- Polgar, L. (1989). *Mechanism of Protease Action*; CRC Press: Boca Raton.
- Powers, A.M., Brault, A.C., Shirako, Y., Strauss, E.G., Kang, W., Strauss, J.H., Weaver, S.C. (2001). Evolutionary relationships and systematics of the alphaviruses. *J. Virol.* 75, 10118-10131.
- Powers, J.C., Asgian, J.L., Ekici, Ö.D. and James, K.E. (2002). Irreversible inhibitors of serine, cysteine, and threonine proteases. *Chemical reviews*, 102(12), pp.4639-4750.
- Preziosi, P. (2007). Isoniazid: metabolic aspects and toxicological correlates. *Current drug metabolism* 8, 839-851.

- Prilusky, J., Felder, C. E., Zeev-Ben-Mordehai, T., Rydberg, E. H., Man, O., Beckmann, J. S., Silman, I. & Sussman, J. L. (2005). FoldIndex©: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* 21, 3435-3438.
- Pushko, P., Parker, M., Ludwig, G. V., Davis, N. L., Johnston, R. E. & Smith, J. F. (1997). Replicon-helper systems from attenuated Venezuelan equine encephalitis virus: expression of heterologous genes in vitro and immunization against heterologous pathogens in vivo. *Virology* 239, 389–401.
- Randolph, J.T., DeGoey, D.A. (2004). Peptidomimetic inhibitors of HIV protease. *Curr. Top. Med. Chem.* 4, 1079-1095.
- Renatus, M., Bode, W., Huber, R., Stuerzebecher, J., & Stubbs, M. T. (1998). Structural and functional analyses of benzamidine-based inhibitors in complex with trypsin: implications for the inhibition of factor Xa, tPA, and urokinase. *Journal of medicinal chemistry*, 41(27), 5445-5456.
- Rey, F. A., Heinz, F. X., Mandl, C., Kunz, C. & Harrison, S. C. (1995). The envelope glycoprotein from tick-borne encephalitis virus at 2 Å resolution. *Nature* 375, 291–298.
- Rezza, G., Nicoletti, L., Angelini, R., Romi, R., Finarelli, A.C., Panning, M., Cordioli, P., Fortuna, C., Boros, S., Magurano, F. and Silvi, G. (2007). Infection with chikungunya virus in Italy: an outbreak in a temperate region. *The Lancet*, 370(9602), pp.1840-1846.
- Rice, C. M. & Strauss, J. H. (1981). Nucleotide sequence of the 26S mRNA of Sindbis virus and deduced sequence of the encoded virus structural proteins. *Proc. Natl. Acad. Sci. USA* 78, 2062–2066.
- Rikkonen, M., Peränen, J. & Kääriäinen, L. (1994). ATPase and GTPase activities associated with Semliki Forest virus non-structural protein nsP2. *J. Virol.* 68, 5804–5810.
- Ross, R.W. (1956). The Newala epidemic: III. The virus: isolation, pathogenic properties and relationship to the epidemic. *J. Hyg. (Lond).* 54, 177-191.
- Rossmann, M. G. & Johnson, J. E. (1989). Icosahedral RNA virus structure. *Ann. Rev. Biochem.* 58, 533–573.
- Rümenapf, T., Brown, D. T., Strauss, E. G., König, M., Rameriz-Mitchel, R. & Strauss, J. H. (1995). Aura alphavirus subgenomic RNA is packaged into virions of two sizes. *J. Virol.* 69, 1741–1746.

Rümenapf, T., Strauss, E. G. & Strauss, J. H. (1994). Subgenomic mRNA of Aura alphavirus is packaged into virions. *J. Virol.* 68, 56–62.

Rümenapf, T., Strauss, E. G. & Strauss, J. H. (1995). Aura virus is a New World representative of Sindbis-like viruses. *Virology* 208, 621–633.

Rupp, J. C., Jundt, N. & Hardy, R. W. (2011). Requirement for the amino-terminal domain of Sindbis virus nsP4 during virus infection. *J. Virol.* 85, 3449–3460.

Rupp, J. C., Sokoloski, K. J., Gebhart, N. N., & Hardy, R. W. (2015). Alphavirus RNA synthesis and non-structural protein functions. *Journal of General Virology*, 96, 2483-2500.

Russo AT, White MA, Watowich SJ (2006) The crystal structure of the Venezuelan equine encephalitis alphavirus nsP2 protease. *Structure* 14, 1449–1458

Ryman, K. D., Gardner, C. L., Burke, C. W., Meier, K. C., Thompson, J. M. & Klimstra, W. B. (2007). Heparan sulfate binding can contribute to the neurovirulence of neuroadapted and nonneuroadapted Sindbis viruses. *J. Virol.* 81, 3563–3573.

Ryman, K. D., Klimstra, W. B. & Johnston, R. E. (2004). Attenuation of Sindbis virus variants incorporating uncleaved PE2 glycoprotein is correlated with attachment to cell-surface heparan sulfate. *Virology* 322, 1–12.

Saha, S., Raghava, G.P.S. (2007). Prediction methods for B-cell epitopes. *Immunoinformatics: Predicting Immunogenicity In Silico*, 387-394.

Sali, A., Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779-815.

Sanz, M. A. & Carrasco, L. (2001). Sindbis virus variant with a deletion in the 6K gene shows defects in glycoprotein processing and trafficking: lack of complementation by a wild-type 6K gene in trans. *J. Virol.* 75, 7778–7784.

Sanz, M. A., Madan, V., Carrasco, L. & Nieva, J. L. (2003). Interfacial domains in Sindbis virus 6K protein. Detection and functional characterization. *J. Biol. Chem* 278, 2051–2057.

Sanz, M. A., Perez, L. & Carrasco, L. (1994). Semliki Forest virus 6K protein modifies membrane permeability after inducible expression in *Escherichia coli* cells. *J. Biol. Chem.* 269, 12106–12110.

Schmaljohn, A. L., & McClain, D. (1996). Alphaviruses (togaviridae) and flaviviruses (flaviviridae) Medical microbiology (4th ed.), University of Texas Medical Branch, Galveston (1996) [Chapter 54]

Schuffenecker, I., Iteaman, I., Michault, A., Murri, S., Frangeul, L., Vaney, M.C., Lavenir, R., Pardigon, N., Reynes, J.M., Pettinelli, F. and Biscornet, L. (2006). Genome microevolution of chikungunya viruses causing the Indian Ocean outbreak. PLoS medicine, 3, p.e263.

Schwartz, O. & Matthew, L. A. (2010). Biology and pathogenesis of chikungunya virus. Nat. Rev. Microbiol. 8, 491–500.

Schwede, T., Kopp, J., Guex, N., Peitsch, M.C. (2003). SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res. 31, 3381-3385.

Sergon K, Njuguna C, Kalani R, Ofula V, Onyango C, Konongoi LS, Bedno S, Burke H, Dumilla AM, Konde J. (2008). Seroprevalence of chikungunya virus (CHIKV) infection on Lamu Island, Kenya, October 2004. The American journal of tropical medicine and hygiene 78, 333-337.

Sergon, K., Yahaya, A.A., Brown, J., Bedja, S.A., Mlindasse, M., Agata, N., Allaranger, Y., Ball, M.D., Powers, A.M., Ofula, V. (2007). Seroprevalence of Chikungunya virus infection on Grande Comore Island, union of the Comoros, 2005. The American journal of tropical medicine and hygiene 76, 1189-1193.

Shah, K. V., Gibbs, C. J. & Banerjee, G. (1964). Virological investigation of the epidemic of haemorrhagic fever in Calcutta: isolation of three strains of chikungunya virus. Indian J. Med. Res. 52, 676–683.

Sharma, R., Fatma, B., Saha, A., Bajpai, S., Sistla, S., Dash, P.K., Parida, M., Kumar, P., Tomar, S. (2016). Inhibition of chikungunya virus by picolinate that targets viral capsid protein. Virology 498, 265-276.

Shirako, Y. & Strauss, J. H. (1994). Regulation of Sindbis virus RNA replication: uncleaved P123 and nsP4 function in minus-strand RNA synthesis, whereas cleaved products from P123 are required for efficient plus-strand RNA synthesis. J. Virol. 68, 1874–1885.

Shirako, Y., Strauss, E. G. & Strauss, J. (2000). 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, 148–160.

Shoemaker, B.A., Zhang, D., Thangudu, R.R., Tyagi, M., Fong, J.H., Marchler-Bauer, A., Bryant, S.H., Madej, T., Panchenko, A.R. (2009). Inferred Biomolecular Interaction Server

web server to analyze and predict protein interacting partners and binding sites. *Nucleic Acids Res.* 38, 518-524.

Singh, I. & Helenius, A. (1992). Role of ribosomes in Semliki Forest virus nucleocapsid uncoating. *J. Virol.* 66, 7049–7058.

Sjöberg, M. & Garoff, H. (2003). Interactions between the transmembrane segments of the alphavirus E1 and E2 proteins play a role in virus budding and fusion. *J. Virol.* 77, 3441–3450.

Skoging, U. & Liljestrom, P. (1998). Role of the C-terminal tryptophan residue for the structure-function of the alphavirus capsid protein. *J. Mol. Biol.* 279, 865–872.

Skoging-Nyberg, U. & Liljeström, P. (2000). A conserved leucine in the cytoplasmic domain of Semliki Forest virus spike protein is important for budding. *Arch. Virol.* 145, 1225–1230.

Smith Iii, A.B., Hirschmann, R., Pasternak, A., Akaishi, R., Guzman, M.C., Jones, D.R., Keenan, T.P., Sprengeler, P.A., Darke, P.L. (1994). Design and synthesis of peptidomimetic inhibitors of HIV-1 protease and renin. Evidence for improved transport. *J. Med. Chem.* 37, 215-218.

Soonsawad, P., Xing, L., Milla, E., Espinoza, J.M., Kawano, M., Marko, M. et al. (2010). Structural evidence of glycoprotein assembly in cellular membrane compartments prior to Alphavirus budding *J. Virol.* 84, 11145–11151.

Soumahoro, M.K., Boelle, P.Y., Gaüzere, B.A., Atsou, K., Pelat, C., Lambert, B., La Ruche, G., Gastellu-Etchegorry, M., Renault, P., Sarazin, M., Yazdanpanah, Y. (2011). The Chikungunya epidemic on La Reunion Island in 2005-2006: a cost-of-illness study. *PLoS neglected tropical diseases* 5, e1197.

Sourisseau, M., Schilte, C., Casartelli, N., Trouillet, C., Guivel-Benhassine, F., Rudnicka, D., Sol-Foulon, N., Le Roux, K., Prevost, M.-C., Fsihi, H. (2007). Characterization of reemerging chikungunya virus. *PLoS Pathog* 3, e89.

Sreerama, N., Woody, R.W. (2000). Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.* 287, 252-260.

Strauss, E. G., & Strauss, J. H. (1986). Structure and replication of the alphavirus genome. In *The Togaviridae and Flaviviridae* (pp. 35-90). Springer New York.

Strauss, E. G., Rice, C. M. & Strauss, J. H. (1984). Complete nucleotide sequence of the genomic RNA of Sindbis virus. *Virology* 133, 92–110.

Strauss, J. H. & Strauss, E. G. (1994). The alphaviruses: gene expression, replication, and evolution. *Microbiol. Rev.* 58, 491–562.

Strauss, J. H. & Strauss, E. G. (2001). Virus evolution: how does an enveloped virus make a regular structure? *Cell* 105, 5–8.

Strauss, J. H., Strauss, E. G. & Kuhn, R. J. (1995). Budding of alphaviruses. *Trends Microbiol.* 3, 346–350.

Strauss, J.H., Strauss, E.G. (1994). The alphaviruses: gene expression, replication, and evolution. *Microbiol. Rev.* 58, 491-562.

Sudo, K., Yamaji, K., Kawamura, K., Nishijima, T., Kojima, N., Aibe, K & Shimizu, Y. (2005). High-throughput screening of low molecular weight NS3-NS4A protease inhibitors using a fluorescence resonance energy transfer substrate. *Antiviral Chemistry and Chemotherapy* 16, 385-392.

Sun, S., Xiang, Y., Akahata, Holdaway, H., W., Pal, P., Zhang, X., Diamond, M. S., Nabel, G. J. & Rossmann, M. G. (2013). Structural analyses at pseudo atomic resolution of Chikungunya virus and antibodies show mechanisms of neutralization. *eLife* 2, e00435.

Suomalainen, M. & Garoff, H. (1992). Alphavirus spike-nucleocapsid interaction and network antibodies. *J. Virol.* 66, 5106–5109.

Suomalainen, M., Liljestrom, P. & Garoff, H. (1992). Spike protein-nucleocapsid interactions drive the budding of alphaviruses. *J. Virol.* 66, 4737–4747.

Tang, J., Jose, J., Chipman, P., Zhang, W., Kuhn, R. J. & Baker, T. S. (2011). Molecular links between the envelope glycoprotein and nucleocapsid core in Sindbis virus. *J. Mol. Biol.* 414, 442–459.

Tellinghuisen TL, Kuhn RJ. (2000). Nucleic acid-dependent cross-linking of the nucleocapsid protein of Sindbis virus. *J. Virol.* 74, 4302–4309

Tellinghuisen TL, Perera R, Kuhn RJ. (2001) *In vitro* assembly of Sindbis virus core-like particles from crosslinked dimers of truncated and mutant capsid proteins. *J. Virol.* 75, 2810–2817

Tellinghuisen, T. L., Hamburger, A. E., Fisher, B. R., Ostendorp, R. & Kuhn, R. J. (1999). *In vitro* assembly of alphavirus cores by using nucleocapsid protein expressed in *Escherichia coli*. *J. Virol.* 73, 5309–5319.

Thiboutot, M.M., Kannan, S., Kawalekar, O.U., Shedlock, D.J., Khan, A.S., Sarangan, G., Srikanth, P., Weiner, D.B., Muthumani, K. (2010). Chikungunya: a potentially emerging epidemic? *PLoS Negl Trop Dis* 4, e623.

Thomas, S., Rai, J., John, L., Günther, S., Drosten, C., Pützer, B. M. & Schaefer, S. (2010). Functional dissection of the alphavirus capsid protease: sequence requirements for activity. *Virology* 407, 327.

Thomas, S., Rai, J., John, L., Schaefer, S., Putzer, B.M. and Herchenröder, O. (2013). Chikungunya virus capsid protein contains nuclear import and export signals. *Virology journal* 10, 269.

Tomar, S., Aggarwal, M. (2017). Structure and function of alphavirus proteases. Academic press, Elsevier (Chapter 5); ISBN: 978-0-12-809712-0.

Tomar, S., Hardy, R.W., Smith, J. L. & Kuhn, R. J. (2006). Catalytic core of alphavirus nonstructural protein nsP4 possesses terminal adenylyltransferase activity. *J. Virol.* 80, 9962–9969.

Tong, L., Wengler, G. & Rossmann, M. G. (1993). Refined structure of sindbis virus core protein and comparison with other chymotrypsin like serine proteinase structures. *J. Mol. Biol.* 230, 228–247.

Uchime, O., Fields, W. & Kielian, M. (2013). The Role of E3 in pH Protection during Alphavirus Assembly and Exit. *J. Virol.* 87, 10255–10262.

Uversky, V. N. (2002). Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* 11, 739–756.

Uversky, V. N., Gillespie, J. R. & Fink, A. L. (2000). Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins: Struct. Funct. Genet.* 41, 415–427.

Uversky, V. N., Oldfield, C. J. & Dunker, A. K. (2008). *Annu. Rev. Biophys.* 37, 215–246.

Vagin, A., Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *Journal of applied crystallography* 30, 1022-1025.

Vasiljeva, L., Merits, A., Auvinen, P. & Kääriäinen, L. (2000). Identification of a novel function of the Alphavirus capping apparatus RNA 5'-triphosphatase activity of Nsp2. *J. Biol. Chem.* 275, 17281–17287.

- Vasiljeva, L., Valmu, L., Kaariainen, L. & Merits, A. (2001). Site-specific protease activity of the carboxyl-terminal domain of Semliki Forest virus replicase protein nsP2. *J. Biol. Chem.* 276, 30786–30793.
- Vihinen H, Ahola T, Tuittila M, Merits A, Kaariainen L. (2001). Elimination of phosphorylation sites of Semliki Forest virus replicase protein nsP3. *J. Biol. Chem.* 276, 5745–5752
- Vihinen H, Saarinen J. (2000). Phosphorylation site analysis of Semliki Forest virus nonstructural protein 3. *J. Biol. Chem.* 275, 27775–27783
- Vogel, R. H., Provencher, S. W., von Bonsdorff, C. H., Adrian, M. & Dubochet, J. (1986). Envelope structure of Semliki Forest virus reconstructed from cryo-electron micrographs. *Nature* 320, 533–535.
- Von Bonsdorff, C. H. & Harrison, S. C. (1975). Sindbis virus glycoproteins form a regular icosahedral surface lattice. *J. Virol.* 16, 141–145.
- Von Bonsdorff, C. H. & Harrison, S. C. (1978). Hexagonal glycoprotein arrays from Sindbis virus membranes. *J. Virol.* 28, 578–583.
- Voss, J. E., Vaney, M. C., Duquerroy, S., Vornrhein, C., Girard-Blanc, C., Crublet, E., et al. (2010). Glycoprotein organization of Chikungunya virus particles revealed by X-ray crystallography. *Nature* 468, 709–712.
- Wahlberg J. & Garoff, H. (1992). Membrane fusion process of Semliki Forest virus I: low pH-induced rearrangement in spike protein quaternary structure precedes virus penetration into cells. *J. Cell Biol.* 116, 339–348.
- Wahlberg, J. M., Boere, W. A. & Garoff, H. (1989). The heterodimeric association between the membrane proteins of Semliki Forest virus changes its sensitivity to low pH during virus maturation. *J. Virol.* 63, 4991–4997.
- Waite MR, Brown DT, Pfefferkorn ER. (1972). Inhibition of Sindbis virus release by media of low ionic strength: an electron microscope study. *J. Virol.* 10, 537–544
- Walker, B. and Lynas, J.F. (2001). Strategies for the inhibition of serine proteases. *Cellular and Molecular Life Sciences*, 58(4), 596-624.
- Wang YF, Sawicki SG, Sawicki DL. (1991). Sindbis virus nsP1 functions in negative-strand RNA synthesis. *J. Virol.* 65, 985–988.

- Wang, E., Volkova, E., Adams, A.P., Forrester, N., Xiao, S.-Y., Frolov, I., Weaver, S.C. (2008). Chimeric alphavirus vaccine candidates for chikungunya. *Vaccine* 26, 5030-5039.
- Wang, Q. M. (1999). Protease inhibitors as potential antiviral agents for the treatment of picornaviral infections. *Prog. Drug Res.* 52, 197–219.
- Warrier R, Linger BR, Golden BL, Kuhn RJ. (2008). Role of Sindbis virus capsid protein region II in nucleocapsid core assembly and encapsidation of genomic RNA. *J. Virol.* 82, 4461–4470
- Weaver, S.C., Winegar, R., Manger, I.D., Forrester, N.L. (2012). Alphaviruses: population genetics and determinants of emergence. *Antiviral Res.* 94, 242-257.
- Weiss, B., Geigenmüller-Gnirke, U. & Schlesinger, S. (1994). Interactions between Sindbis virus RNAs and a 68 amino acid derivative of the viral capsid protein further defines the capsid binding site. *Nucl. Acids Res.* 22, 780–786.
- Weiss, B., Nitschko, H., Ghattas, I., Wright, R. & Schlesinger, S. (1989). Evidence for specificity in the encapsidation of Sindbis virus RNAs. *J. Virol.* 63, 5310–5318.
- Wengler, G., Boege, U., Bischoff, H. & Wahn, K. (1982). The core protein of the alphavirus Sindbis virus assembles into core-like nucleoproteins with the viral genome RNA and with other single-stranded nucleic acids in vitro. *Virology* 118, 401–410.
- Wengler, G., Koschinski, A. & Dreyer, F. (2003) Entry of alphaviruses at the plasma membrane converts the viral surface proteins into an ion-permeable pore that can be detected by electrophysiological analyses of whole-cell membrane currents. *J. Gen. Virol.* 84, 173–181.
- Wengler, G., Würkner, D. & Wengler, G. (1992). Identification of a sequence element in the alphavirus core protein which mediates interaction of cores with ribosomes and the disassembly of cores. *Virology* 191, 880–888.
- White, J., Kartenbeck, J. & Helenius, A. (1980). Fusion of Semliki forest virus with the plasma membrane can be induced by low pH. *J. Cell Biol.* 87, 264–272.
- Wiederstein, M., Sippl, M.J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 35, 407-410.
- Wilkinson, T. A., Tellinghuisen, T. L., Kuhn, R. J. & Post, C. B. (2005). Association of sindbis virus capsid protein with phospholipid membranes and the E2 glycoprotein: implications for alphavirus assembly. *Biochemistry* 44, 2800–2810.

Wirth, D. F., Katz, F., Small, B. & Lodish, H. F. (1977). How a single Sindbis virus mRNA directs the synthesis of one soluble protein and two integral membrane glycoproteins. *Cell* 10, 253–263.

Wu, S. R., Haag, L., Sjoberg, M., Garoff, H. & Hammar, L. (2008). The dynamic envelope of a fusion class II virus. E3 domain of glycoprotein E2 precursor in Semliki Forest virus provides a unique contact with the fusion protein E1. *J. Biol. Chem.* 283, 26452–26460.

Wu, Y., Eigenbrot, C., Liang, W.C., Stawicki, S., Shia, S., Fan, B., Ganesan, R., Lipari, M.T. and Kirchhofer, D. (2007). Structural insight into distinct mechanisms of protease inhibition by antibodies. *Proceedings of the National Academy of Sciences*, 104, 19784–19789.

Yadav, M., Dubey, M. L., Gupta, I., Bhatti, G. & Malla, N. (2007). Cysteine proteinase 30 in clinical isolates of *T. vaginalis* from symptomatic and asymptomatic infected women. *Exp. Parasitol.* 116, 399–406

Yadav, M., Dubey, M. L., Gupta, I. & Malla, N. (2007). Cysteine proteinase 30 (CP30) and antibody response to CP30 in serum and vaginal washes of symptomatic and asymptomatic *Trichomonas vaginalis* infected women. *Parasite immunol.* 29, 359–365.

Yadav, V. K., Chhikara, N., Gill, K., Dey, S., Singh, S., & Yadav, S. (2013). Three low molecular weight cysteine proteinase inhibitors of human seminal fluid: purification and enzyme kinetic properties. *Biochimie*, 95(8), 1552-1559.

Yadav, V. K., Mandal, R. S., Puniya, B. L., Singh, S., & Yadav, S. (2015). Studies on the interactions of SAP-1 (an N-terminal truncated form of cystatin S) with its binding partners by CD-spectroscopic and molecular docking methods. *Journal of Biomolecular Structure and Dynamics*, 33(1), 147-157.

Yamane, J., Yao, M., Zhou, Y., Hiramatsu, Y., Fujiwara, K., Yamaguchi, T., Yamaguchi, H., Togame, H., Tsujishita, H., Takemoto, H. and Tanaka, I. (2011). In-crystal affinity ranking of fragment hit compounds reveals a relationship with their inhibitory activities. *Journal of Applied Crystallography*, 44, 798-804.

Yang, W., Wang, L., & Paschen, W. (2013). Development of a high-throughput screening assay for inhibitors of small ubiquitin-like modifier proteases. *Journal of biomolecular screening*, 18, 621-628.

Yao, J. S., Strauss, E. G. & Strauss, J. H. (1996). Interactions between PE2, E1, and 6K required for assembly of alphaviruses studied with chimeric viruses. *J. Virol.* 70, 7910–7920.

Zhang, R., Hryc, C. F., Cong, Y., Liu, X., Jakana, J., Gorchakov, R., Baker, M. L., Weaver, S. C. & Chiu, W. (2011). 4.4 Å cryo-EM structure of an enveloped alphavirus Venezuelan equine encephalitis virus. *EMBO J.* 30, 3854–3863.

Zhang, W., Fisher, B. R., Olson, N. H., Strauss, J. H., Kuhn, R. J. & Baker, T. S. (2002). Aura virus structure suggests that the T=4 organization is a fundamental property of viral structural proteins. *J. Virol.* 76, 7239–7246.

Zhang, W., Mukhopadhyay, S., Pletnev, S. V., Baker, T. S., Kuhn, R. J. & Rossmann, M. G. (2002). Placement of the structural proteins in Sindbis virus. *J. Virol.* 76, 11645–11658.

Zhao, H. & Garoff, H. (1992). Role of cell surface spikes in alphavirus budding. *J. Virol.* 66, 7089–7095.

Zhao, H., Lindqvist, B., Garoff, H., von Bonsdorff, C. & Liljeström, P. (1994). A tyrosine-based motif in the cytoplasmic domain of the alphavirus envelope protein is essential for budding. *EMBO J.* 13, 4204–4211.

