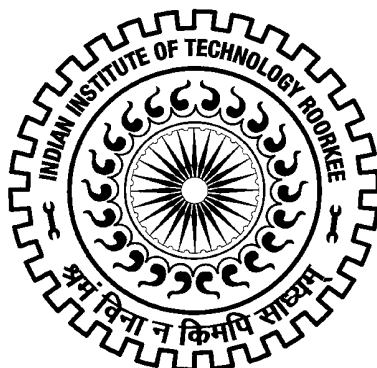


# MOLECULAR CHARACTERIZATION OF AURA VIRUS CAPSID PROTEIN

Ph.D. THESIS

*by*

MEGHA AGGARWAL



DEPARTMENT OF BIOTECHNOLOGY  
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE  
ROORKEE – 247667 (INDIA)  
SEPTEMBER, 2014



**MOLECULAR CHARACTERIZATION OF  
AURA VIRUS CAPSID PROTEIN**

**A THESIS**

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

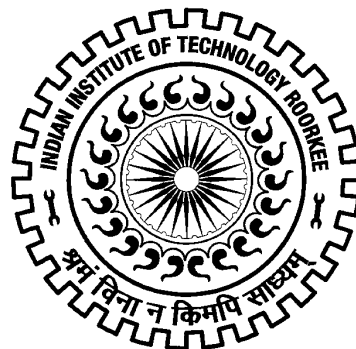
**DOCTOR OF PHILOSOPHY**

*in*

**BIOTECHNOLOGY**

*by*

**MEGHA AGGARWAL**



**DEPARTMENT OF BIOTECHNOLOGY  
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE  
ROORKEE – 247667 (INDIA)  
SEPTEMBER, 2014**

**© INDIAN INSTITUTE OF TECHNOLOGY ROORKEE-2014  
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, **“MOLECULAR CHARACTERIZATION OF AURA 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 is an authentic record of my own work carried out during a period from September 2009 to September 2014 under the supervision of Dr. Shailly Tomar, Associate Professor, Department of Biotechnology, Indian Institute of Technology Roorkee.

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

(MEGHA AGGARWAL)

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

(Shailly Tomar)  
Supervisor

Date:



## **Abstract**

Alphaviruses, the members of *Togaviridae* family, are enveloped, arthropod borne, single-stranded RNA viruses. The alphavirus genus includes Aura virus, Sindbis virus, Chikungunya virus (CHIKV), Semliki Forest virus and Venezuelan equine encephalitis virus (VEEV). The recent outbreaks of Chikungunya epidemic pose a threat to human health and hence, the development of an effective antiviral drug to combat alphavirus infections has become a high priority. We have structurally and biochemically investigated the properties of capsid protein (CP) from Aura and Chikungunya virus for the development of novel strategies to identify inhibitors against alphavirus infection.

The CP is a multifunctional protein and performs several functions including structural polyprotein processing, genomic RNA encapsidation, nucleocapsid formation and virus budding for the alphavirus propagation. The CP interacts with the E2 glycoprotein cytoplasmic tail which leads to virus budding. In order to analyze the interaction strategy, the CP from Aura virus (AVCP) has been cloned, expressed and purified to homogeneity. The crystal structure of the AVCP is determined. Also, the homology models of the cytoplasmic tails of glycoproteins were generated. The crystal structure of AVCP and the molecular models of E1 and E2 cytoplasmic tails are fitted into the cryo-electron microscopic (cryo-EM) density map of VEEV. The fitting studies reveal the presence of the helix-loop-helix motif in the cytoplasmic tail and the loop region of the helix-loop-helix motif is predicted to bind the hydrophobic pocket of the CP. The disulfide bridge, ionic and hydrophobic interactions between the two opposite helices stabilize the motif. The conserved Pro405 residue in the loop region of the helix-loop-helix motif of E2 tail binds and makes direct molecular interactions in the hydrophobic pocket. The crystal structure of AVCP in complex with dioxane has been determined and molecular contacts of the ligand with the protein have been analyzed. The dioxane molecule binds to the CP hydrophobic pocket to which E2 tail binds and is found to be present exactly at the same position where Pro405 is predicted to bind to the pocket. Molecular docking studies for various dioxane based compounds were performed and docked structures were analyzed for studying the binding mode of potential inhibitory dioxane derivatives. Additionally, another crystal structure of AVCP in complex with piperazine also shows the similar binding pattern of the ligand to the hydrophobic pocket. Interestingly, structural analysis of AVCP in complex with picolinic acid identifies a ligand binding site on the surface of CP, which is separate from the hydrophobic pocket.

The alphavirus CP possessing serine protease activity after cleaving itself off from the rest of the polyprotein by *cis*-proteolysis becomes enzymatically inactive for the rest of the virus life cycle. Structural analysis of the *trans*-proteolytic active form of AVCP has been performed. The C-terminal truncated AVCP (last two residues including Trp267 deleted) has been cloned, expressed and purified. The crystal structure of the active AVCP has been determined and compared with various forms (inactive and substrate bound inactive) of the CP from different alphaviruses. The comparative structural analysis shows the differences in the C-terminal residues which are missing in the active form, highly flexible in substrate bound form and remains intact in Trp bound post-cleavage inactive form. The active site, oxyanion hole and the specificity pockets show marked variations in different enzymatic forms. The hydrophobic pocket of the different forms of the protein also shows conformational switching. Additionally, the *trans*-protease activity has been performed using the fluorogenic peptide substrate and the kinetic parameters were determined. The FRET (fluorescence resonance energy transfer) based *trans*-protease activity assay has also been optimized for CHIKV capsid protease. The assay can be developed further for high throughput inhibitor screening against alphavirus infection. Thus, the structural details and the *in vitro* FRET based protease assay provides the worthwhile guidelines for the structure based designing and testing of the protease inhibitors of the CP from alphavirus which will lead to the inhibition of the first step of structural polyprotein processing.

The thesis consists of five chapters which include the structural and biochemical characterization of the CP from Aura virus, a member of the genus alphavirus. The molecular characterization of the CP involves the 3D structural analysis, *in silico* modeling, cryo-EM fitting and biochemical studies. The *trans*-proteolytic enzyme activity of alphavirus CP has been determined and characterized from Aura virus and CHIKV.

**Chapter 1** reviews the literature. It describes the alphavirus life cycle, its transmission and the structure of the overall virion. The alphavirus genome organization and the polyprotein processing including both structural and non-structural polyprotein have been described in detail. The nucleocapsid assembly and the virus budding process have also been depicted. The chapter also describes the overall structure of CP and its involvement in the process of virus budding, genomic RNA encapsidation and nucleocapsid assembly. The CP has been studied as a proteolytic enzyme and different inhibition strategies for alphavirus capsid protease are discussed in the chapter.



**Chapter 2** describes the structural analysis of the AVCP and insight into the capsid-glycoprotein interaction. The recombinant CP has been produced in *E. coli* and purified to homogeneity by affinity and size exclusion chromatography methods. The crystallization of the purified protein was performed and the crystal structure was determined. The crystal structure reveals the presence of chymotrypsin-like structure having Greek key motif. Subsequently, the homology models of E1 and E2 glycoproteins were made. The crystal structure of AVCP and the homology models of E1 and E2 glycoproteins cytoplasmic tails were fitted into the cryo-EM structure of VEEV. The fitted structure was analyzed for the molecular interactions of the capsid and glycoproteins. The structure shows the presence of helix-loop-helix motif of which the loop region binds to the hydrophobic pocket of the CP. The Cys-Cys disulfide bridge present in between the two helices has been proposed to stabilize this helix-loop-helix motif. Likewise, the Tyr401 ionic and hydrophobic interactions with the opposite helix also seem to be stabilizing the motif. The loop residue Pro405 binds directly to the hydrophobic pocket and might be involved in the capsid-glycoprotein interaction during virus budding process.

**Chapter 3** describes the structural analysis of AVCP in complex with potential CP inhibitory molecules. The crystal structure of AVCP in complex with dioxane, piperazine and picolinic acid was determined and analyzed for the ligand binding studies. For preparation of the ligand bound crystals, the native protein crystals were soaked into the reservoir buffer containing various ligands. The structural analysis of complexes reveals the presence of dioxane in the hydrophobic pocket of the CP. This dioxane-bound structure was fitted into the cryo-EM structure of VEEV along with the homology models of E1 and E2 cytoplasmic tails as described in chapter 2. The dioxane molecule was found to be present exactly at the same position where Pro405 of E2 tail binds to the pocket. Thus, it can be concluded that the dioxane molecule mimics the pyrrolidine ring of the proline and might compete with E2 tail for the binding to the hydrophobic pocket. The piperazine also binds at the same position in the hydrophobic pocket where the dioxane binds and shows similar interactions with the pocket residues. Interestingly, the picolinic acid binds on the surface of the capsid surrounded by the neighboring symmetry related molecules. Bound picolinic acid also shows interactions with the symmetry CP molecules as well. Other potential antiviral compounds based on dioxane molecule have been designed which contain two dioxane moieties. The docking of these compounds shows that the two ring structures bind into two separate grooves of the CP hydrophobic pocket.

**Chapter 4** demonstrates the structural insight into the active form of the capsid protease from Aura virus and the comparative studies with the different forms of the enzyme. As the Trp267 binds to the specificity pocket after the *cis*-autoproteolytic cleavage and inhibits *trans*-protease activity by blocking the binding of substrate to the specificity pocket. Therefore, the last two residues including Trp267 were deleted from the capsid construct to activate the capsid protease by unblocking the active site. The C-terminal truncated construct was cloned into the expression vector, expressed in *E. coli* and purified using the affinity and gel filtration chromatography. The crystals of the active truncated form were produced and the crystal structure was determined. The structure was compared with the other enzymatic forms revealing the differences in the C-terminal loop after His261, the specificity pockets and the active site residues. The oxyanion hole consists of a conserved water molecule which is absent in the post-cleavage form where water molecule is replaced by the scissile bond residues. The hydrophobic pocket also shows the differences in between the active and inactive forms stating the conformational switching in the pocket from binding with the other capsid monomers to interaction with the cytoplasmic tail of E2 glycoprotein. This indicates that the conformational changes in the hydrophobic pocket can lead to switching of the CP function from nucleocapsid formation to virus budding. The different specificity pockets show conformational variations in the active and inactive state. Thus, the chapter describes the 3D structure of active CP that has been previously proposed to be natively unfolded protein.

**Chapter 5** reports the *trans*-proteolytic activity of alphavirus capsid protease. The AVCP was characterized for its *trans*-proteolytic activity using the fluorogenic peptide substrate having EDANS (fluorophore) and DABCYL (quencher) at the ends. The peptide sequence that separates these tags includes the sequence derived from the scissile bond of the capsid protease. After the cleavage of the peptide in presence of active CP, the fluorescence intensity is expected to increase due to the disruption of FRET (Fluorescence resonance energy transfer). Kinetic parameters using fluorogenic peptide substrates for the Aura virus capsid protease were estimated; and a  $K_M$  value was found to be  $2.63 \pm 0.62 \mu\text{M}$ , and a  $K_{\text{cat}}/K_M$  value was  $4.97 \times 10^4 \text{ M}^{-1} \text{ min}^{-1}$ . Additionally, for the activity analysis of the CHIKV CP (CVCP), cloning, expression and purification of CVCP to homogeneity was performed. The  $K_M$  and  $K_{\text{cat}}/K_M$  values of purified CVCP were determined using fluorogenic peptide substrate and were found to be  $1.27 \pm 0.34 \mu\text{M}$  and  $5.5 \times 10^4 \text{ M}^{-1} \text{ min}^{-1}$  respectively. The CVCP was characterized for the optimization of pH and NaCl concentration. The

effect of glycerol on the enzyme activity has also been determined. In conclusion, this chapter explains the development of high throughput screening method for proteolytic activity assay of alphavirus capsid protease that can be used for screening of alphavirus specific protease inhibitors.



## ACKNOWLEDGEMENTS

The present thesis work is due to the effort from a number of people. Foremost, I express my sincere gratitude towards my supervisor, Dr. Shailly Tomar. Since I began my graduate studies, I am enlightened by her passion for science, excellent guidance and developing me as a scientist. Without her encouragement, the thesis work would not have been possible. Dr. Shailly Tomar provided me all the resources to ensure the successful voyage of my Ph. D. Her emphasis on critical thinking and reading along with the hard work at the bench influenced me a lot. Her unlimited energy, patience and trust in me are very inspiring. I feel very fortunate to have been a part of her research group. She was always available whenever I needed her during my thesis work. Ever since, she has supported me academically and emotionally through the rough roads to complete my thesis.

The valuable scientific studies are always the result of collaborative efforts, and I enjoyed working with the folks both inside and outside the Institute. I endow my special thanks to Dr. Pravindra Kumar, for helping me out in the structural studies. I couldn't have succeeded without his immense support in structure biology and bioinformatics. I would like to thank him for critical scientific and non-scientific discussions. He always encouraged me to perform better. He appreciated my work all the time during my thesis which boosted me up for more and more hard work. I must also thank Dr. Richard Kuhn and Dr. M. M. Parida for providing the helpful guidance and resources to accomplish my thesis work.

My research committee members guided me through all these years. I am very obliged to my committee members Dr. Partha Roy, Dr. R. P. Singh and Dr. R. K. Peddinti for providing me the feedback on my research work. The advice of the members is highly appreciated. I am very fortunate to get the opportunity of having such great committee members.

I am grateful to Prof. Ritu Barthwal, Coordinator, NMR facility, for her help and support in accomplishing the experiments involving fluorescence studies. It is my great pleasure to thank the faculty members Prof. G. S. Randhawa, Dr. R. Prasad, Dr. Vikas Pruthi, Dr. Bijan Chaudhary, Dr. Sanjoy Ghosh, Dr. A. K. Sharma, Dr. Naveen K. Navani, Dr. Ranjana Pathania, Dr. Maya Nair for their help and encouragement.

I owe my sincere thanks to Dr. Manali Datta, Dr. Preeti, Dr. Dipak Patil, Dr. Satya Tapas, Dr. Shivendra Pratap, Dr. Aditya Dev, Dr. Pramod Kumar and Ms. Sakshi Tomar to be instrumental for helping me in acclimate to the laboratory. They represented me numerous

experimental methods with myriad of techniques that were essential for my thesis work. I had a great time in laboratory with Dr. Preeti and Dr. Satya who helped me a lot as lab mates and friends as well. Ms. Sakshi helped me regarding the protease activity and structural analysis included in this thesis.

I thank Dr. Girijesh Patel, Mr. Prabhat Tomar, Ms. Sonali Dhindwal, Ms. Manju Narwal, Mr. Bibekananda Kar, Ms. Nidhi Sharma, Ms. Anamika Singh for the fruitful discussions. I had immense pleasure working with these people and got support in various experiments included into my thesis. I appreciate the discussions with Dr. Selva Kumar who helped me whenever I was not getting positive results during my research work. I also thank Mr. Rajat Mudgal, Mr. Rajesh Sharma and Ms. Ramanjit Kaur specially to facilitate the fun environment in the laboratory and also for helping me in all the possible ways during my tenure.

I would like to thank Ms. Preeti Verma, Mrs. Monu Batra, Mr. Madhusudan Rao, Mr. Harvijay Singh, Mr. Vijay Sharma, Ms. Anchal Sharma, Ms. Neha Singh, Ms. Gunjan Saini, Ms. Pooja Kesari, Ms. Anjali Malik and Mr. Pranav for their sincere cooperation and efforts. I have also had the pleasure to work with Ms. Benazir Fatma, Mr. Ashutosh Gupta, Ms. Vishnupriya Pandey and Mr. Rameez Mehar. They contributed well for the accomplishment of the studies included in the thesis.

I must also thank Mr. Pradeep TP, Ms. Sweta Tripathi and Dr. Swati Srivastava. As colleagues and friends, they encouraged me throughout my thesis time period. This thesis would not exist without their love and support. Through all my good and bad times during my thesis, these people were with me all the time and helped me in every way. I am heartiest thankful to Mr. Pradeep and Ms. Sweta for helping me in the fluorescence experiments included in this thesis. I greatly value their friendship and deeply appreciate their care that helped me to stay focused on my studies. Without them I would not be able to fill colors into my thesis.

I am also grateful to the Institute Sports Club as well as the Institute Hospital. Both helped me in maintaining physical fitness so that I could accomplish my studies contained in this thesis. I had a great time with Ms. Nidhi Sharma and Ms. Ramanjit kaur who encouraged me to utilize the sports facility. I also appreciate the help from The Department Library. I am also thankful to the departmental staff who maintained the department instruments so efficiently.

Finally, and most importantly, my heartiest thank goes to my parents and all of my family members for their blessings and unconditional love. They always encourage me to pursue 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 almighty God for blessing me and keeping me strong and calm during all good and hard times throughout my life.

Finally, I would like to thank Council of Scientific & Industrial Research (CSIR), New Delhi India for financial assistance as JRF and SRF.

(Megha Aggarwal)





## LIST OF PUBLICATIONS

1. **Megha Aggarwal**, Sonali Dhindwal, Shivendra Pratap, Richard J. Kuhn, Pravindra Kumar and Shailly Tomar. Crystallization, high-resolution data collection and preliminary crystallographic analysis of the Aura virus capsid protease and its complex with Dioxane. *Acta Cryst.* (2011). F67, 1394–1398
2. **Megha Aggarwal**, Satya Tapas, Preeti, Anjul Siwach, Pravindra Kumar, Richard J. Kuhn and Shailly Tomar. Crystal structure of aura virus capsid protease and its complex with dioxane: new insights into capsid–glycoprotein molecular contacts. *PLoS One* 7, e51288 (2012). DOI: 10.1371/journal.pone.0051288.
3. **Megha Aggarwal**, Sonali Dhindwal, Pravindra Kumar, Richard J. Kuhn and Shailly Tomar *trans*-Protease activity and structural insights into the active form of alphavirus capsid protease. *J. Virol.* (2014). DOI:10.1128/JVI.01692-14.
4. **Megha Aggarwal**, Rajesh Sharma Pravindra Kumar, Manmohan Parida and Shailly Tomar. Analysis of *trans*-enzymatic activity of Chikungunya virus capsid protease and development of FRET-based high throughput inhibitor screening assay. (In preparation)
5. Monu Batra, Rajesh Sharma, Sonali Dhindwal, **Megha Aggarwal**, Pawan Gupta, Rajesh Pratap Singh, Shailly Tomar. *In silico* structural and proteomic analysis of *Bacillus subtilis* chemotaxis methyltransferase CheR reveal SAH binding and regulatory role of N-terminal domain. (In preparation)



## CONTENTS

Page No.

### CANDIDATE'S DECLARATION

ABSTRACT.....	i-v
ACKNOWLEDGEMENTS.....	vii-ix
LIST OF PUBLICATIONS.....	xi
CONTENTS.....	xiii-xvii
LIST OF FIGURES.....	xix-xxiii
LIST OF TABLES.....	xxv
LIST OF ABBREVIATIONS USED.....	xxvii-xxix
STRUCTURES SUBMITTED TO PROTEIN DATA BANK.....	xxxi
INTRODUCTION.....	1-3

## CHAPTER 1

### REVIEW OF LITERATURE

1.1 Introduction.....	5
1.2 Alphavirus life cycle.....	5
1.3 The alphavirus virion.....	10
1.4 The alphavirus genome.....	10
1.5 Alphavirus Non-structural proteins.....	11
1.5.1 Non-structural polyprotein processing.....	11
1.5.2 Functions of Non-structural proteins.....	12
1.6 Alphavirus Structural proteins.....	13
1.6.1 Structural polyprotein processing.....	14
1.6.2 Functions of structural proteins.....	15
1.7 The nucleocapsid.....	17
1.8 Virus budding.....	19
1.8.1 Role of Glycoproteins in viral budding.....	20
1.8.2 Role of 6K in viral budding.....	21
1.8.3 Role of virus RNA in budding.....	21

1.9 Structure of alphavirus capsid protein.....	22
1.10 Role of Capsid protein in Virus budding.....	25
1.11 RNA encapsidation and capsid assembly.....	26
1.12 Viral proteases.....	28
1.12.1 Capsid protein as serine protease.....	30
1.12.1.1 The catalytic triad of alphavirus CP.....	31
1.12.1.2 Oxyanion hole.....	32
1.12.1.3 Specificity pockets of alphavirus CP .....	32
1.12.1.4 Trans-activity of capsid protein.....	34
1.13 Strategies to inhibit alphavirus capsid protein.....	35
1.13.1 Protein-protein interaction inhibition through dioxane analogs.....	35
1.13.2 Capsid protease inhibition through the substrate-based inhibitors.....	37

## **CHAPTER 2**

### **CRYSTAL STRUCTURE OF AURA VIRUS CAPSID PROTEIN AND INSIGHT INTO THE CAPSID-GLYCOPROTEIN INTERACTION**

2.1 Abstract.....	39
2.2 Introduction.....	40
2.3 Materials and Methods.....	42
2.3.1 Materials.....	42
2.3.2 Strategy used for the study.....	42
2.3.2.1 Construction of expression plasmids.....	44
2.3.2.2 Expression of the capsid protein.....	45
2.3.2.3 Purification of AVCP.....	45
2.3.2.4 Crystallization.....	46
2.3.2.5 Data collection.....	46
2.3.2.6 Structure determination and refinement.....	46
2.3.2.7 Molecular modeling.....	47
2.4 Results and Discussion.....	47
2.4.1 Prediction of the disordered region in AVCP.....	48
2.4.2 Cloning and heterologous expression of the capsid protein.....	48
2.4.3 Purification of AVCP.....	50

2.4.4 Crystal development and data collection.....	51
2.4.5 Overall structure of AVCP.....	53
2.4.6 Comparative analysis of AVCP with capsid proteases from different alphaviruses...57	
2.4.7 Homology Modeling of E1 and E2 glycoproteins.....	59
2.4.8 Interaction analysis of E1 and E2 glycoproteins with the capsid protease.....	60
2.5 Conclusion.....	67

### **CHAPTER 3**

#### **ASSESSMENT OF THE COMPLEXES OF AVCP WITH DIOXANE AND SIMILAR COMPOUNDS**

3.1 Abstract.....	69
3.2 Introduction.....	70
3.3 Methodology.....	72
3.3.1 Strategy used for the study.....	72
3.3.2 Co-Crystallization and Data Collection.....	72
3.3.3 Structure solution and refinement.....	73
3.3.4 Molecular Docking and Analysis.....	73
3.4 Results and Discussion.....	73
3.4.1 Crystal structure of the dioxane bound AVCP.....	74
3.4.2 Comparison of the dioxane bound complex structure with the apo form of AVCP.....	77
3.4.3 Inhibition strategy to disrupt capsid and glycoprotein interaction.....	79
3.4.4 Complex of AVCP with piperazine.....	81
3.4.5 Complex of AVCP with picolinic acid.....	83
3.4.6 Docking analysis.....	88
3.4.6.1 Docking of dioxane based and similar compounds.....	88
3.4.6.2 Binding of the compounds to the hydrophobic pocket.....	91
3.4.6.3 Interaction of the dioxane and other compounds to the hydrophobic pocket.....	91
3.5 Conclusion.....	95

## **CHAPTER 4**

### **STRUCTURAL INSIGHTS INTO THE ACTIVE FORM OF THE ALPHAVIRUS CAPSID PROTEASE**

4.1 Abstract.....	97
4.2 Introduction.....	97
4.3 Methodology.....	100
4.3.1 Construction of expression plasmid.....	100
4.3.2 Expression and Purification of AVCP $\Delta$ 2.....	101
4.3.3 Crystallization and data collection.....	102
4.3.4 Structure solution and refinement.....	102
4.4 Results and Discussion.....	103
4.4.1 Purification of active AVCP $\Delta$ 2.....	103
4.4.2 Crystallization of AVCP $\Delta$ 2.....	105
4.4.3 Overall structure of AVCP $\Delta$ 2.....	107
4.4.4 Dimeric interactions.....	113
4.4.5 Conformational switching in the hydrophobic pocket.....	114
4.4.6 The catalytic triad.....	117
4.4.7 The oxyanion hole.....	119
4.4.8 The substrate specificity pockets.....	121
4.5 Conclusion.....	128

## **CHAPTER 5**

### **TRANS-PROTEOLYTIC ACTIVITY OF ALPHAVIRUS CAPSID PROTEASE**

5.1 Abstract.....	131
5.2 Introduction.....	132
5.3 Materials and methods.....	136
5.3.1 Molecular cloning and Construction of expression plasmid.....	136
5.3.2 Expression of both inactive and active CVCP.....	137
5.3.3 Purification of CVCP (inactive and active).....	138
5.3.4 trans-Protease activity assay.....	138
5.3.6 Effect of pH and NaCl concentration on enzyme activity.....	139
5.3.7 Effect of Glycerol.....	139

5.4 Results and Discussion.....	140
5.4.1 Expression and Purification of recombinant CVCP (both active and inactive forms) in E. coli.....	140
5.4.2 <i>In vitro</i> trans-protease assay of CVCP.....	142
5.4.3 Effect of buffer pH on enzyme activity.....	144
5.4.4 Effect of NaCl concentration on enzyme activity.....	144
5.4.5 Glycerol tolerance.....	145
5.4.6 Increasing enzyme concentration verses proteolytic activity.....	146
5.4.7 Inhibition studies.....	146
5.4.8 Characterization of trans-protease activity for active form of AVCP.....	147
5.5 Conclusion.....	149
<b>REFERENCES.....</b>	<b>153-170</b>





## LIST OF FIGURES

	Page No.
<b>Figure 1.2.1:</b> Schematic representation of the alphavirus transmission cycle.....	6
<b>Figure 1.2.2:</b> Overview of alphavirus life cycle.....	7
<b>Figure 1.2.3:</b> Schematic representation of heterodimerization of E1 and E2 glycoproteins.....	8
<b>Figure 1.2.4:</b> Structural alignment of the mature E3-E2-E1 proteins and immature p62-E1 proteins.....	9
<b>Figure 1.4.1:</b> Alphavirus genome organization.....	11
<b>Figure 1.5.1.1:</b> Processing of non-structural polyprotein.....	12
<b>Figure 1.6.1.1:</b> Processing of Structural polyprotein.....	14
<b>Figure 1.7.1:</b> Schematic representation of the CP.....	18
<b>Figure 1.8.1:</b> Schematic of two types of budding models.....	20
<b>Figure 1.9.1:</b> The crystal structures of the CP from different alphaviruses.....	24
<b>Figure 1.10.1:</b> Hydrophobic pocket of SINV CP with bound dioxane.....	26
<b>Figure 1.11.1:</b> The N-terminal arm of the CP binds to the hydrophobic pocket of the neighboring CP which mimics the interaction of E2 glycoprotein with CP....	28
<b>Figure 1.12.1:</b> Catalytic mechanism of the serine proteases.....	30
<b>Figure 1.12.1.1.1:</b> Active site of SCP.....	31
<b>Figure 2.2.1:</b> Cartoon view of the cryo-EM structure of VEEV (PDB ID: 3J0C).....	41
<b>Figure 2.3.2.1:</b> Schematic representation of the methodology used for the study of AVCP.....	43
<b>Figure 2.4.1.1:</b> Structurally disordered region prediction in AVCP.....	48
<b>Figure 2.4.2.1:</b> The PCR amplification of the full length AVCP (left panel). The right panel shows the restriction enzyme digestion of the cloned plasmid confirming the presence of correct clone.....	49
<b>Figure 2.4.2.2:</b> Expression of the full length AVCP.....	49
<b>Figure 2.4.2.3:</b> Agarose gel electrophoresis of the PCR amplified product of capsid protease domain of Aura virus (left panel). The cloned plasmid was screened for the presence of right construct and the restriction enzyme digestion shows the digested band at 500 bp which confirms the positive cloned plasmid (right panel).....	50
<b>Figure 2.4.2.4:</b> SDS-PAGE gel for the over-expression of AVCP.....	50
<b>Figure 2.4.3.1:</b> Gel-filtration profile and SDS-PAGE analysis of AVCP. ....	51

<b>Figure 2.4.4.1:</b> Crystals of Aura virus capsid protease (left). Diffraction of AVCP crystals using in-house radiation (right panel) .....	51
<b>Figure 2.4.5.1:</b> Structure of Aura Virus Capsid Protease.....	54
<b>Figure 2.4.5.2:</b> Multiple sequence alignment of AVCP with CPs from other alphaviruses.....	55
<b>Figure 2.4.5.3:</b> The surface view of the crystal structure of AVCP showing the hydrophobic pocket in red.....	57
<b>Figure 2.4.6.1:</b> Structural alignment of CPs from different alphaviruses.....	58
<b>Figure 2.4.7.1:</b> The cartoon representation of the E2 glycoprotein showing the ectodomain as well as endodomain.....	59
<b>Figure 2.4.7.2:</b> The homology model of the E1 glycoprotein is shown in the cartoon view.....	60
<b>Figure 2.4.8.1:</b> Representation of the crystal structure of AVCP-dioxane complex and the homology models of E1 and E2 glycoproteins fitted into the cryo-EM electron density map of VEEV.....	61
<b>Figure 2.4.8.2:</b> The cytoplasmic domain of E2 glycoprotein forms the helix-loop-helix motif. The loop of this motif interacts with the hydrophobic pocket of the CP. The E1 protein also comes close to the capsid which might be involved in CP-glycoprotein interaction. ....	61
<b>Figure 2.4.8.3:</b> Cartoon view of the AVCP-glycoproteins interaction.....	62
<b>Figure 2.4.8.4:</b> The cytoplasmic tail of E2 glycoprotein from Aura virus showing the Cys-Cys disulfide bond in between helix-loop-helix motif of E2.....	63
<b>Figure 2.4.8.5:</b> The structural alignment of the cytoplasmic domain of E2 glycoproteins from AURAV and VEEV.....	63
<b>Figure 2.4.8.6:</b> Multiple sequence alignment of the cytoplasmic domain of E2 (cdE2) from different alphaviruses.....	64
<b>Figure 2.4.8.7:</b> Structural alignment of cdE2 predicted structure from Aura virus (pink) and VEEV (blue).....	65
<b>Figure 2.4.8.8:</b> The critical polar interactions between E2 and capsid as well as within the helix-loop-helix motif of cdE2 are shown in dotted lines.....	66
<b>Figure 3.2.1:</b> Hydrophobic pocket of different chains of SCP (PDB ID: 1WYK).....	71
<b>Figure 3.3.1.1:</b> The schematic representation of the strategy used for the study.....	72
<b>Figure 3.4.1:</b> The structure of the compounds dioxane, piperazine and picolinic acid.....	74

<b>Figure 3.4.1.1:</b> AVCP–dioxane complex crystal diffraction.....	74
<b>Figure 3.4.1.2:</b> AVCP-dioxane complex structure showing the dioxane binding pocket.....	75
<b>Figure 3.4.1.3:</b> Surface view of the dioxane (blue) bound AVCP.....	76
<b>Figure 3.4.1.4:</b> The hydrophobic pocket of the AVCP complexed with dioxane as revealed by crystal structural studies.....	76
<b>Figure 3.4.1.5:</b> Superimposition of dioxane bound forms of AVCP (violet) and SCP (green) crystal structures highlighting the hydrophobic pocket residues.....	77
<b>Figure 3.4.2.1:</b> Structural comparison of the hydrophobic pocket in both apo and dioxane bound complex form of AVCP.....	78
<b>Figure 3.4.2.2:</b> The structural alignment of the hydrophobic pocket forming residues from apo (red) and dioxane bound form (blue) of AVCP.....	79
<b>Figure 3.4.3.1:</b> The dioxane molecule (blue) in the crystal structure of AVCP-dioxane complex occupies the same position as Pro405 (pink) present in the helix-loop-helix motif of cdE2. ....	80
<b>Figure 3.4.4.1:</b> The diffraction pattern of AVCP-piperazine complex crystal.....	81
<b>Figure 3.4.4.2:</b> The surface and stick view of the AVCP in complex with the piperazine.....	82
<b>Figure 3.4.4.3:</b> Comparison of the apo, dioxane bound and piperazine bound AVCP.....	83
<b>Figure 3.4.5.1:</b> The diffraction pattern of AVCP-picolinic acid complex crystal.....	84
<b>Figure 3.4.5.2:</b> The interaction of the picolinic acid at the surface of AVCP.....	84
<b>Figure 3.4.5.3:</b> The interaction strategy of picolinic acid to AVCP.....	86
<b>Figure 3.4.6.1.1:</b> Compounds used for the docking analysis based on previous studies.....	89
<b>Figure. 3.4.6.1.2:</b> Compounds used for docking analysis which were derived from PubChem based on the structure of dioxane.....	90
<b>Figure 3.4.6.2.1:</b> The binding of four different compounds (1, 2, 3, 13) into the hydrophobic pocket of the AVCP.....	91
<b>Figure 3.4.6.3.1:</b> The dioxane molecule present in the hydrophobic pocket shows the hydrophobic interactions with the pocket residues.....	92
<b>Figure 3.4.6.3.2:</b> The interaction of the compound 1 with the hydrophobic pocket residues of the AVCP .....	93
<b>Figure 4.2.1:</b> Multiple sequence alignment of the residues at scissile bond.....	98
<b>Figure 4.4.1.1:</b> Cloning of AVCP $\Delta$ 2 construct.....	103
<b>Figure 4.4.1.2:</b> Expression of AVCP $\Delta$ 2 construct.....	104

<b>Figure 4.4.1.3:</b> Gel filtration chromatography and SDS-PAGE analysis of major peak fractions show the protein (AVCP $\Delta$ 2) purified to homogeneity.....	104
<b>Figure 4.4.2.1:</b> Crystals obtained for the active form of AVCP and the diffraction pattern.....	105
<b>Figure 4.4.3.1:</b> Overall structure of active AVCP $\Delta$ 2 construct.....	108
<b>Figure 4.4.3.2:</b> Overall structure of the monomer contains two $\beta$ -barrel subdomains consisting of the catalytic triad in between the cleft.....	108
<b>Figure 4.4.3.3:</b> Surface view of AVCP $\Delta$ 2 showing different pockets and regions involved in catalysis, coded with different colors.....	109
<b>Figure 4.4.3.4:</b> The S <sub>1</sub> specificity pocket of SCP substrate bound mutant form.....	110
<b>Figure 4.4.3.5:</b> Both chains of active and substrate-bound form were aligned along with native AVCP.....	111
<b>Figure 4.4.3.6:</b> The superposition of the active and native AVCP shows the presence of one glycerol molecule (blue in color) exactly at the same position where Trp267 (red in color) side chain lies in the specificity pocket.....	112
<b>Figure 4.4.3.7:</b> The interaction of the glycerol with different residues is shown.....	112
<b>Figure 4.4.4.1:</b> Close-up view of the dimeric interaction in AVCP $\Delta$ 2.....	114
<b>Figure 4.4.5.1:</b> The N-terminus of CP bound to the hydrophobic pocket of neighboring CP molecule .....	115
<b>Figure 4.4.5.2:</b> The surface view and the superposed view of the hydrophobic pocket of AVCP $\Delta$ 2 over AVCP crystal structure.....	117
<b>Figure 4.4.6.1:</b> The catalytic triad residues from different enzymatic states: native AVCP (pink), AVCP $\Delta$ 2 Chain A (green) and AVCP $\Delta$ 2 Chain B (yellow) show conformational change in Ser218 side chain.....	118
<b>Figure 4.4.7.1:</b> The differences at the S <sub>1</sub> specificity pocket of AVCP $\Delta$ 2 are shown in superposition with native AVCPcrystal structure with bound Trp267 at the active site.....	119
<b>Figure 4.4.7.2:</b> The differences in the volume of S <sub>1</sub> specificity pocket.....	120
<b>Figure 4.4.7.3:</b> Oxyanion hole of AVCP $\Delta$ 2 (green) and Chain A of substrate bound SCP (blue) occupy a water molecule. However, the H-bonding occurs between oxyanion hole residues and carbonyl oxygen atom preceding scissile bond in chain B of substrate bound form and native AVCP.....	121

<b>Figure 4.4.8.1:</b> Schematic representation of different specificity pockets interacting with different substrate residues.....	122
<b>Figure 4.4.8.2:</b> S <sub>1</sub> Specificity pocket of AVCPΔ2 chain A (green) and chain B (yellow) is compared with that of native AVCP (pink).....	123
<b>Figure 4.4.8.3:</b> Surface view of S <sub>1</sub> specificity pocket of all three enzymatic forms.....	124
<b>Figure 4.4.8.4:</b> Leu234 main chain forms H-bond with Ser218 as well as Trp267.....	125
<b>Figure 4.4.8.5:</b> Multiple Sequence alignment of alphavirus CPs and chymotrypsin.....	126
<b>Figure 4.4.8.6:</b> Comparison of the S <sub>4</sub> pocket.....	127
<b>Figure 4.4.8.7:</b> S <sub>4</sub> ' and S <sub>2</sub> ' pockets comparison.....	128
<b>Figure 5.2.1:</b> Schematic representation of inactive and active forms of alphavirus CP.....	132
<b>Figure 5.2.2:</b> Fluorescence resonance energy transfer (FRET) assay.....	135
<b>Figure 5.4.1.1a:</b> PCR amplification of structural polyprotein from Chikungunya virus;	
<b>b:</b> Cloning of inactive and active CVCP.....	140
<b>Figure 5.4.1.2:</b> SDS-PAGE analysis of expression and solubility of recombinant Chikungunya inactive and active protein.....	141
<b>Figure 5.4.1.3:</b> SDS-PAGE and gel filtration profile of both inactive (upper panel) and active (lower panel) CVCP.....	141
<b>Figure 5.4.2.1:</b> <i>In vitro</i> trans-proteolytic assay for CVCP.....	142
<b>Figure 5.4.2.2:</b> Kinetic studies of CVCP.....	143
<b>Figure 5.4.3.1:</b> The effect of pH on the enzymatic activity was observed.....	144
<b>Figure 5.4.4.1:</b> The effect of NaCl concentration on the enzymatic activity was observed.....	145
<b>Figure 5.4.5.1:</b> Dependence of the CVCP proteolytic activity on the glycerol concentration.....	145
<b>Figure 5.4.6.1:</b> Effect of enzyme concentration on the proteolytic activity of the CVCP.....	146
<b>Figure 5.4.8.1:</b> Analysis of <i>In vitro trans</i> -proteolytic activity of AVCPΔ2.....	147
<b>Figure 5.4.8.2:</b> Kinetic studies of active AVCP.....	148
<b>Figure 5.4.8.3:</b> The influence of the buffer pH on the proteolytic activity of the active AVCP has been observed.....	149



## LIST OF TABLES

	Page No.
<b>Table 1.1:</b> Functions of Non-Structural proteins.....	13
<b>Table 1.2:</b> Functions of Alphavirus Structural proteins.....	15
<b>Table 1.3:</b> CP structural data from different members of alphavirus family.....	23
<b>Table 1.4:</b> Residues involved in the formation of different specificity pockets.....	33
<b>Table 1.5:</b> Dioxane analogs and their antiviral activity.....	36-37
<b>Table 2.1:</b> Data collection and refinement statistics for AVCP in apo form.....	52
<b>Table 3.1:</b> Data collection and refinement statistics for AVCP in complex with different ligands.....	86-87
<b>Table 3.2:</b> The docking studies with the AVCP for the dioxane analogs.....	94
<b>Table 3.3:</b> The Glide Score and Glide Emodel of the compounds derived from PubChem according to their rank for docking score.....	95
<b>Table 4.1:</b> Data collection and refinement statistics for active form of AVCP.....	106
<b>Table 5.1:</b> List of the oligonucleotides used.....	137





## LIST OF ABBREVIATIONS USED

Å	Angstrom
$\alpha$	Alpha
°C	Degree Centigrade
$\beta$	Beta
$\mu\text{g}$	Microgram
$\mu\text{l}$	Microlitre
$\mu\text{M}$	Micromolar
$\mu\text{mole}$	Micromole
$\gamma$	Gamma
AVCP	Aura virus capsid protein
BamH	Bacillus amyloliquefaciens
bp	Base pair
B-factor	Debye-Waller factor/Temperature factor
CD	Circular dichroism
cDNA	complementary DNA
CHIKV	Chikungunya virus
cm	Centimeter
CP	Capsid Protein
CPV	Cytopathic vacuole
CSE	Conserved sequence element
Da	Daltons
DNA	Deoxyribose nucleic acid
dNTPs	Deoxy nucleotide tri phosphates
E. coli	Escherichia coli
EDTA	Ethylenediaminetetracetic acid
eg.	For example
<i>et al.</i>	et alia
FRET	Fluorescence resonance energy transfer
g	Gram
h	Hours

H-bond	Hydrogen bond
IITR	Indian Institute of Technology Roorkee
IPTG	Isopropyl $\beta$ -D-thiogalactoside
K	Kelvin
kbp	Kilo bases
kDa	Kilo Daltons
kV	kiloVolt
L.	Liter
m	Meter
M	Molar
mA	milliampere
mg	milligram
min	Minute
ml	millilitre
mm	millimeter
mM	millimolar
M-MuLV	Moloney Murine Leukemia virus
NaCl	Sodium Chloride
NCBI	National Center for Biotechnology Information
Nde	Neisseria denitrificans
Ni-NTA	Nickel-nitrilotriacetic acid
nm	Nanometer
NMR	Nuclear magnetic resonance
Nsp	Non-structural protein
PCR	Polymerase chain reaction
PDB	Protein Data Bank
RE	Restriction endonuclease
R-factor	Residual-factor
RNA	Ribose nucleic acid
RNasin	RNase Inhibitor
rpm	Revolutions per minute
s	Seconds

SCP	Sindbis virus capsid protein
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SFV	Semliki Forest virus
T	Triangulation number
USA	United States of America
v/v	volume /volume
VEEV	Venezuelan equine encephalitis virus
w/v	weight /volume
Xho	Xanthomonas holcicola



## **STRUCTURES SUBMITTED TO PROTEIN DATA BANK**

PDB ID: 4AGJ

PDB ID: 4AGK

PDB ID: 4UON

Piperazine bound AVCP complex

Picolinic acid bound AVCP complex



**Introduction**

The alphaviruses belong to the *Togaviridae* family of viruses. They are single stranded, positive sense RNA viruses having 11.5 kb long genome. The alphaviruses are enveloped, arthropod borne viruses that cause a myriad of diseases from mild fever, skin rashes to harsh encephalitis and arthritis. The members of the genus include well studied Sindbis virus, Semliki Forest virus, Eastern, Western and Venezuelan equine encephalitis virus, Aura virus and Chikungunya virus. The members of the genus are divided into two categories, the old world alphaviruses and the new world alphaviruses; according to the place of their origin and disease transmission. Recent threats of Chikungunya make it essential to work on alphaviruses and to develop the new strategies for the inhibition of alphavirus infection.

The cryo-electron microscopic structures of the different alphaviruses are available which provide us insight into the structure and organization of the virion. The alphaviruses has T=4 icosahedral symmetry for both glycoproteins arrangement and capsid assembly. The genomic RNA is encapsidated by the capsid protein (CP) to form the nucleocapsid assembly which is surrounded by the lipid bilayer. The glycoproteins are arranged on the lipid bilayer in icosahedral symmetry.

The alphavirus genome consists of the non-structural proteins coding gene in its two-third N-terminus region while the C-terminus one-third region encodes for the structural proteins. The genomic 49S RNA translates into the non-structural polyprotein and also transcribes the 26S sub-genomic RNA via minus strand RNA intermediate. The 26S sub-genomic RNA translates into the structural polyprotein. The structural polyprotein consists of the CP at the amino-terminus followed by the E3, E2, 6K and E1 proteins. The glycoproteins form the virus envelope and function for virus release, entry and disassembly. The CP performs multiple functions in the virus life cycle. It has auto-proteolytic activity and thus cleaves itself from rest of the polyprotein and helps in structural polyprotein processing. It interacts with the genomic RNA to encapsidate the genome. The CP interacts with other CP monomers for the formation of capsid assembly. The CP also interacts with the E1 and E2 glycoproteins leading to virus budding. The CP consists of two major domains: the amino terminal domain and the carboxyl terminal domain. The N-terminus consists of highly basic residues and interacts with the genomic RNA to encapsidate it.

The carboxyl terminal domain is a chymotrypsin like serine protease. It acts as cis-protease and cleaves itself between CP and E3 glycoprotein. The scissile bond contains the WS residues, of which W is the last tryptophan residue of CP and S is the initial serine residue in E3 glycoprotein.

## Introduction

After the auto-catalytic cleavage, the tryptophan residue remains bound to the specificity pocket near active site and prevents further protease action of the CP. Thus, CP performs the proteolytic action just once in the virus life cycle. The C-terminal domain also contains the hydrophobic pocket for the binding of glycoproteins which leads to virus release. The structure of the CP from Sindbis virus revealed the basic structure of the CP that is similar to the chymotrypsin. The inactive form of the CP has only been crystallized till date in which the last tryptophan residue remains bound to the active site.

In the present study, we have investigated new strategies for the development of antivirals against alphaviruses. The CP from Aura virus has been studied for its interaction mechanism with the viral glycoproteins that leads to virus budding and its protease activity which is the first essential proteolytic activity required for the processing of the structural polyprotein. The crystal structure of Aura virus CP has been determined and the structural analysis was performed. In the absence of the E1 and E2 glycoproteins crystal structures, the homology models of the cytoplasmic domain of both the proteins were generated. The crystal structure and homology models of AVCP and glycoproteins respectively have been fitted into the cryo-EM density map of Venezuelan equine encephalitis virus (VEEV). Analysis of the cryo-EM fitted 3D structures reveal details of the molecular interactions likely to occur between nucleocapsid and glycoproteins during the virus budding process. The hydrophobic pocket of CP is involved in the interactions with the glycoproteins. The structures of the ligand bound complexes to the hydrophobic pocket of AVCP have also been determined and examined. Dioxane molecule binds to the pocket where the cytoplasmic domain of E2 glycoprotein (cdE2) is supposed to make the molecular contacts. Thus, this suggests that the dioxane derivative compounds might act as competitive protein binding inhibitors for disrupting the binding of the cdE2 protein to the pocket, halting the virus budding process. Based on dioxane, some other dioxane based compounds were also searched from the database. These were used for the *in silico* docking analysis and in CP-complex crystal structure analysis. Thus, the dioxane, its derivatives and other dioxane based compounds were established as the inhibitors for the CP-glycoprotein interaction and the virus budding.

The proteolytically active form of the AVCP has also been crystallized and the structure has been determined. The overall chymotrypsin like fold of active AVCP is same as that of the inactive protein, however, there are some conformational variations in the active site, oxyanion hole and specificity pockets. The comparative studies for the active form were performed with the inactive Aura virus CP and the substrate bound inactive form of Sindbis virus CP. An *in vitro*



FRET (Fluorescence resonance energy transfer) based assay has been established for the first time to determine the *trans*-protease activity of alphavirus CP. The protease activity has been characterized for the CP from Aura virus and Chikungunya virus. The FRET based proteolytic assay can be used for developing a high throughput assay for screening inhibitors against the alphavirus capsid protease.



## 1.1 Introduction

Alphaviruses are members of the *Togaviridae* family, possessing single-stranded positive-sense genomic RNA. They are arthropod borne, enveloped viruses and worldwide in distribution (Strauss & Strauss, 1994). They are the causative agents of diseases ranging from mild fever to harsh encephalitis and may also be a menace of bioterrorism (Reichert *et al.*, 2009; Sherman & Griffin, 1990). The members of this genus include Chikungunya virus (CHIKV), Eastern, Western and Venezuelan equine encephalitis virus, Aura virus, Semliki Forest virus (SFV) and Sindbis virus (SINV). At present, there is no human vaccine or medical treatment available against alphavirus infection.

Aura virus was first isolated from *Aedes serratus* in Brazil and northern Argentina (Rümenapf, Strauss *et al.*, 1995). Although the members of the alphavirus family share a high protein sequence identity of greater than 36 %, they interestingly exhibit differences in RNA packaging, virus assembly, inhibition of host cellular processes and specific infectivity. Unlike other alphaviruses, Aura virus capsid protein (CP) encapsidates both the 4.2 kb subgenomic RNA as well as the genomic RNA in virion particles. This is a unique feature of Aura virus which is responsible for their less precise virus assembly in comparison to other alphaviruses (Rümenapf *et al.*, 1994). In recent years, Chikungunya virus, an arthritogenic alphavirus, has been considered as an emerging threat to human health, which may lead to epidemics (Simon *et al.*, 2011; Hoarau *et al.*, 2010; Schwartz & Matthew, 2010).

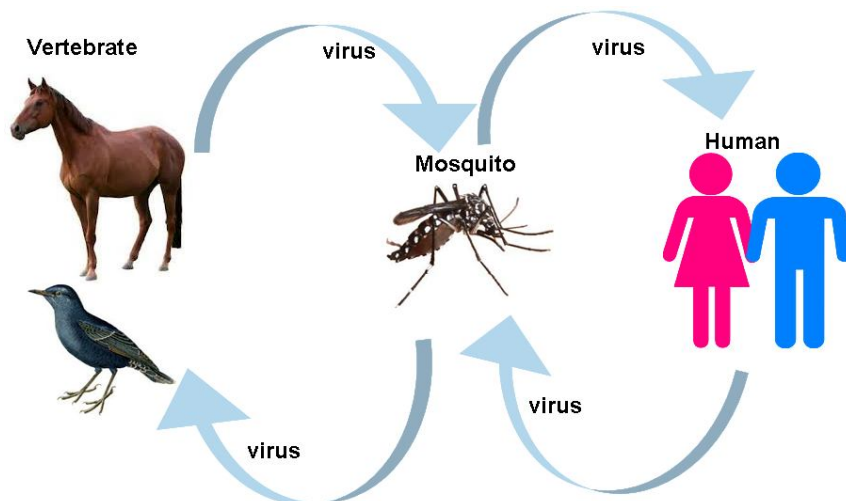
## 1.2 Alphavirus life cycle

The alphavirus transmission in nature occurs through mosquito-vertebrate-mosquito cycles. The schematic diagram of the alphavirus transmission is presented in Figure 1.2.1. The infected mosquitoes transmit the alphavirus infection into the vertebrate hosts. Birds are the common primary vertebrate hosts. The infection begins when a mosquito bites a viremic host and acquires the viral contagion. The viral replication takes place inside the mosquito and transmitted further through the salivary secretion of the vector. Humans can get the infection either through mosquito bite or through other vertebrate hosts.

The life cycle begins with the attachment of virus particle to the host cell membrane using specific host cell receptors. Alphaviruses utilize a wide variety of cell receptors from different tissues and species, thus can infect a range of species from mammals to mosquitoes to humans

## Chapter 1: Review of literature

(Kielian, 1995). These receptors include cell surface heparin sulphate, laminin receptors,  $\alpha 1\beta 1$  integrin and C-type lectins DC-SIGN and L-SIGN (Klimstra *et al.*, 2003; 1998).

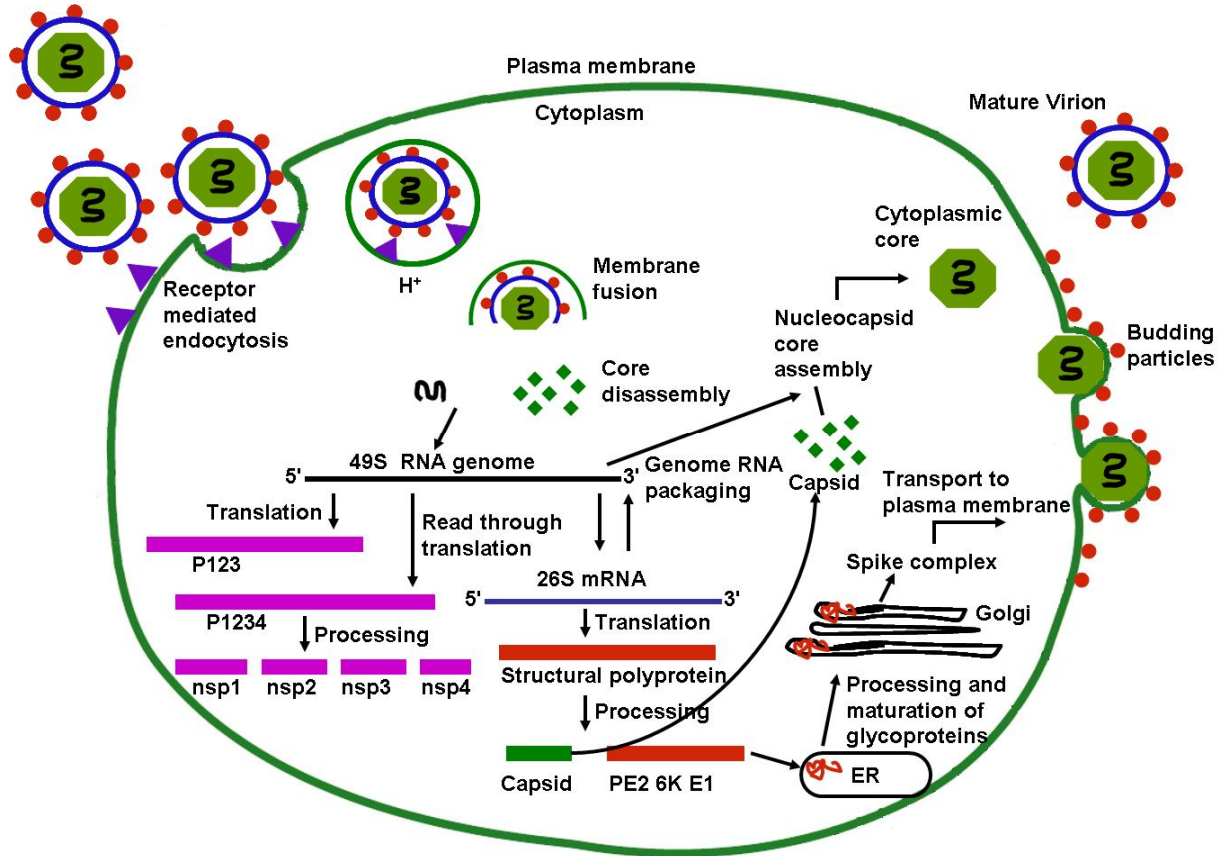


**Figure 1.2.1:** Schematic representation of the alphavirus transmission cycle.

The simplified schematic representation of alphavirus life cycle is given in Figure 1.2.2. The first step of virus attachment is followed by receptor mediated endocytosis (DeTulleo & Kirchhausen, 1998; Helenius *et al.*, 1980). In endosomes, due to the presence of the ATP dependent proton pumps (vacuolar ATPase) in the endosomal membrane, acidification takes place that leads to fusion of the virion particle to the endosomal membrane (Justman *et al.*, 1993; Wahlberg & Garoff, 1992; Mellman *et al.*, 1986; Kielian and Helenius, 1985). The fusion to the endosomal membrane is carried out due to the E1-E2 dimer dissociation followed by conformational changes in viral glycoprotein E1 (Kondor-Koch *et al.*, 1983). Afterwards, the nucleocapsid core is released into the cytoplasm which interacts with ribosomes and the genomic RNA gets freed (Singh & Helenius, 1992). The genomic RNA serves as an mRNA and translates into non-structural polyprotein (P123 and P1234) (Strauss *et al.*, 1984). The two different polyproteins are generated by read through of a stop codon in polyprotein (Li & Rice, 1989). After polyprotein processing, four non-structural proteins nsP1, nsP2, nsP3 and nsP4 are produced. These non-structural proteins along with some unknown factors form replication complex.

The genomic RNA gives rise to 26S sub-genomic RNA via a minus strand RNA intermediate. The subgenomic RNA gets translated into the structural polyprotein which gets processed to produce CP, E3, E2, 6K and E1 glycoproteins. The CP possesses autoproteolytic activity and it gets cleaved itself off from rest of the polyprotein. It encapsidates genomic RNA to

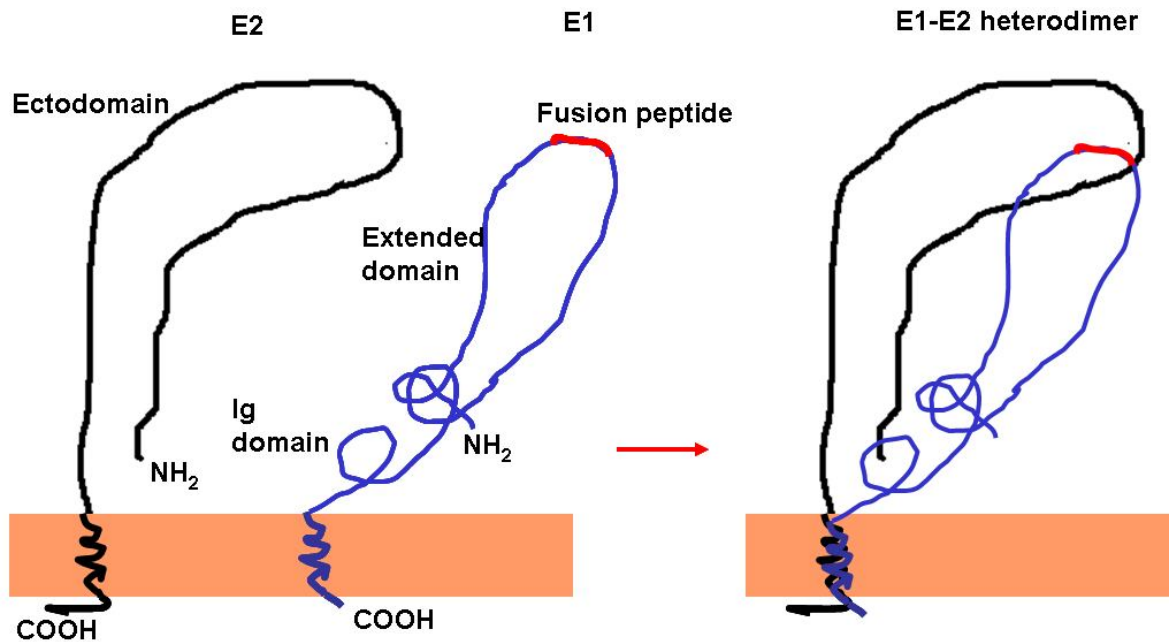
form nucleocapsid assembly. The glycoproteins are transported through endoplasmic reticulum and golgi bodies to the plasma membrane.



**Figure 1.2.2: Overview of alphavirus life cycle.** The virus particle enters the cell via receptor mediated endocytosis. After encountering the low pH in endosome, the conformational rearrangements in the glycoproteins take place which results in core disassembly. The released genomic RNA acts as an mRNA to translate the non-structural polyprotein which further gets processed to generate four non-structural proteins. The genomic RNA transcribed into 26S subgenomic RNA through minus strand RNA intermediate. The 26S subgenomic RNA translates to form structural polyprotein at the amino-terminus of which the CP is present. The CP cleaves itself from rest of the polyprotein which translocates to endoplasmic reticulum and golgi for further processing. The CP encapsidates genomic RNA to form nucleocapsid assembly and interacts with the glycoproteins that leads to budding of the virus particle.

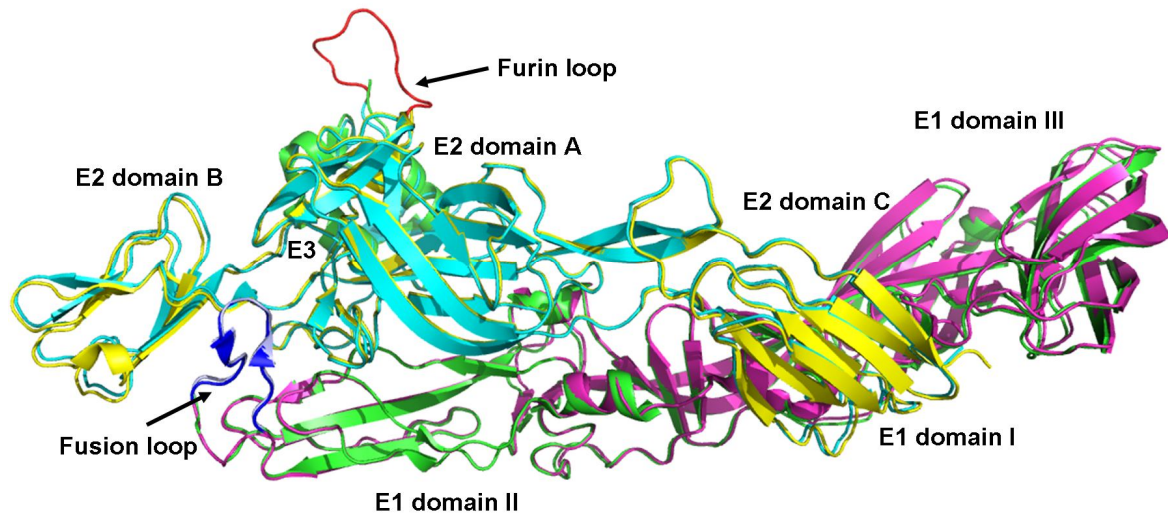
In the endoplasmic reticulum, E1 and E2 form glycoprotein heterodimer. When it comes in contact with the acidic pH of endosomes, the dissociation of the dimer takes place and the rearrangement of the glycoprotein E1 occurs which further acts as fusion protein (Li *et al.*, 2010; Wahlberg *et al.*, 1989; Marsh *et al.*, 1983; White & Helenius, 1980; White *et al.*, 1980). At neutral pH, the fusion peptide hides behind the E1-E2 heterodimer which gets exposed at acidic pH and helps in fusion with the endosomal membrane. The schematic representation of the E1 and E2 heterodimerization is shown in Figure 1.2.3. The acidic pH and the disruption of the host membrane by gamma peptides (virus encoded lytic peptides) are also required for the entry of the

non-enveloped viruses inside the host cell (Odegard *et al.*, 2009; Banerjee *et al.*, 2009; Johnson & Banerjee, 2008).



**Figure 1.2.3: Schematic representation of heterodimerization of E1 and E2 glycoproteins.** E1 protein consists of the fusion peptide along with the other domains. The heterodimerization of E1 and E2 glycoproteins takes place before the furin cleavage. The fusion peptide is masked behind the E1-E2 heterodimer at neutral pH. When it meets the acidic pH, the conformational changes occur and the fusion peptide gets exposed to help in virus fusion to the endosomal membrane.

E1 and E2 glycoproteins consist of ecto and endo domains. The cytoplasmic domain of glycoprotein interacts with the CP that leads to budding of alphaviruses and the generation of mature virions. The crystal structures of immature p62-E1 heterodimer and the mature E3-E2-E1 after furin cleavage have been reported (Voss *et al.*, 2010). E1 interacts with E2 glycoprotein laterally through its domain II. E3 protein seems to stabilize the relative orientation of E2 domain A with respect to B. It also consists of the groove at which the fusion loop of E1 binds, confirming the role of E3 protein in E1-E2 heterodimerization (Lobigs *et al.*, 1990). The structures of the mature and immature virion proteins are similar except for the absence of furin loop in the mature form. This loop consists of the furin site which gets disordered upon cleavage. The crystal structure alignment of both mature E3-E2-E1 and immature p62-E1 heterodimer is shown in Figure 1.2.4. The furin loop is visible only in the immature form while the rest of the domains and the fusion loop remain unchanged.



**Figure 1.2.4: Structural alignment of the mature E3-E2-E1 proteins and immature p62-E1 proteins.** The structural comparison shows the similar structure for both the forms except the absence of furin loop in the mature form which gets disordered after cleavage. The furin loop is shown in red. Fusion loop is in blue. Different domains of E1, E2 and E3 proteins are colored differently. (PDB IDs: 3N42, 3N40)

In the infected cells, the translation of the viral proteins directs the synthesis of cytopathic structures. These structures are formed due to the reorganization of cellular membranes in association with virus specific components (Froshauer *et al.*, 1988; Grimley *et al.*, 1972; Acheson & Tamm, 1967). The electron microscopic analysis of these cytopathic vacuoles (CPVs) have been performed in SFV and identified as CPV-I and CPV-II (Friedman *et al.*, 1972; Grimley *et al.*, 1968). The CPV-I has been found to derive from the modification of endosomal and lysosomal membranes, however, CPV-II are trans-golgi derived (Kujala *et al.*, 2001; Griffiths *et al.*, 1989; Griffiths *et al.*, 1983). CPV-I forms very early in infection (after ~ 1 to 2 hrs of infection) and acts as the site for RNA replication (Grimley *et al.*, 1968; Acheson & Tamm, 1967). These structures are found to be present in wild type as well as mutant virus infected cells (Zhao *et al.*, 1994).

In SFV, CPV-II is found very late in the infection and is less characterized as compared to the CPV-I (Kujala *et al.*, 2001; Barton *et al.*, 1991). This CPV-II has been observed to be involved in the alphavirus budding as the budding process is halted in the absence of this structure (Garoff *et al.*, 1994; Zhao *et al.*, 1994). The electron tomography and immuno-electron microscopy studies have been performed to find out the role of CPV-II in the virus life cycle (Soonsawad *et al.*, 2010). The studies demonstrated the presence of an array of E1/E2 glycoproteins which are arranged helically. Thus, the role of CPV-II in cellular trafficking of the glycoproteins has been observed which explains its function in viral budding.

### 1.3 The alphavirus virion

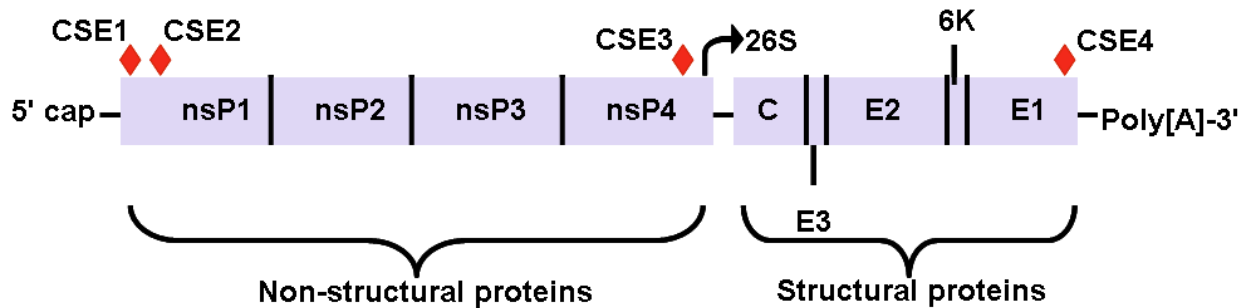
Alphavirus virions are enveloped with icosahedral symmetry having a diameter of 70 nm. The cryo-electron microscopy and image reconstruction techniques have been used to determine alphavirus structure (Pletnev *et al.*, 2001; Mancini *et al.*, 2000; Cheng *et al.*, 1995). The cryo-electron microscopic (cryo-EM) structures for SINV, Ross River virus (RRV), SFV, Aura virus, Venezuelan equine encephalitis virus (VEEV), CHIKV are available at high resolutions (Sun *et al.*, 2013; Zhang *et al.*, 2011; Mukhopadhyay *et al.*, 2006; Zhang *et al.*, 2002; Mancini *et al.*, 2000; Cheng *et al.*, 1995). The overall structural features for all members of the alphavirus family are very similar to each other. The molecular weight and density of a virus particle are  $5.2 \times 10^6$  Da and  $1.22 \text{ g/cm}^3$  respectively (Paredes *et al.*, 1992).

The virion consists of 80 glycoprotein spikes which penetrate through the host derived lipid bilayer (Zhang *et al.*, 2002). Each spike is made up of heterodimer of E1 and E2 glycoproteins that are assembled into trimers. These spikes are arranged in T=4 icosahedral symmetry (Paredes *et al.*, 1993; Fuller, 1987; Vogel *et al.*, 1986; von Bonsdorff & Harrison, 1975). A complete virion consists of 240 copies of individual E2 and E1 glycoproteins. Both the glycoproteins E1 and E2 are traversed through the virus membrane and the cytoplasmic domain of E2 interacts with the nucleocapsid. The nucleocapsid is also arranged in T=4 icosahedral symmetry (Mukhopadhyay *et al.*, 2002). Also, 240 copies of nucleocapsid form the nucleocapsid assembly. CP encapsidates the viral genomic RNA to form nucleocapsid cores. However, in Aura virus, the subgenomic RNA is also encapsidated by the CP (Rümenapf *et al.*, 1995).

### 1.4 The alphavirus genome

The alphavirus genome is positive-sense- single stranded RNA of 11.7 Kb (Kuhn, 2007). It comprises of 5' cap and Poly-A tail at the 3' end. The genomic RNA includes two open reading frames (ORFs): the Non-structural ORF and the Structural ORF. Non-structural ORF (N-terminus two third part of genome) encodes for the non-structural proteins which are involved in various functions of virus life cycle such as replication, polyprotein processing, RNA transcription and capping. The structural ORF (C-terminus one-third part of genome) encodes for structural polyprotein that includes capsid and glycoproteins. The genome organization of alphavirus is shown in Figure 1.4.1.





**Figure 1.4.1: Alphavirus genome organization.** The genomic RNA consists of 5' cap and 3' poly-A tail. The N-terminus two third region of the alphavirus genome encodes for the non-structural proteins nsP1, nsP2, nsP3 and nsP4. The C-terminus one third region encodes for the structural polyprotein which includes the amino terminus CP and E1, E2, E3 glycoproteins. The four CSEs present in the genome helps in the transcription of the genomic RNA. CSE3 present just upstream to the 26S RNA initiation site, is involved in the transcription of subgenomic RNA. Other CSEs also work for minus strand RNA synthesis and the enhancement of the viral replication.

The genome has four conserved sequence elements (CSEs) which interact with viral or host proteins and modulate alphavirus replication. The first CSE (CSE1) is present at the 5' end of the genome and consists of 44 nucleotides. The region in minus strand RNA that is complementary to CSE1 acts as promoter for sub-genomic RNA synthesis (Ou *et al.*, 1982). The second CSE (CSE2) is 51 nucleotides long and present at the 5' end which is probably involved in minus strand RNA synthesis (Frolov *et al.*, 2001). In the presence of 24 nucleotides long CSE3 at the articulation of the structural and non-structural proteins, subgenomic RNA transcription is very efficient (Pushko *et al.*, 1997). CSE4 contains 19 nucleotides and is located at the 3' region of the genome and again affects minus strand RNA synthesis as it is believed to initiate the contact of 5' and 3' ends of genomic RNA (Hardy, 2006; Kuhn *et al.*, 1990).

## 1.5 Alphavirus Non-Structural proteins

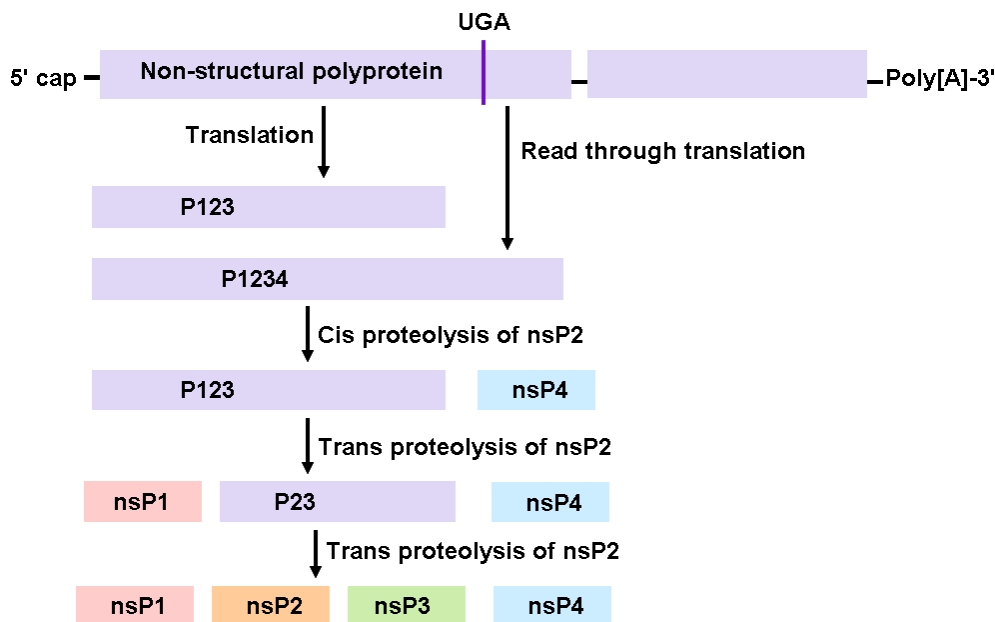
The two third N-terminal part of the genomic RNA is translated to form the non-structural proteins. All the non-structural proteins are required for genomic RNA replication and sub-genomic RNA transcription.

### 1.5.1 Non-Structural polyprotein processing

The schematic representation of polyprotein processing is shown in Figure 1.5.1.1. The genomic RNA is translated into two polyproteins (P123 and P1234). Due to the presence of an opal codon UGA between nsP3 and nsP4, polyprotein P123 is produced. However, 10-20 % times read through of the opal codon leads to the formation of P1234 (Shirako & Strauss, 1994; Li &

## Chapter 1: Review of literature

Rice, 1993; de Groot *et al.*, 1990; Li & Rice, 1989).



**Figure 1.5.1.1: Processing of non-structural polyprotein.** The N-terminus two third part of genomic RNA encodes for non-structural polyprotein. The opal codon UGA is present in the coding region. P123 is translated by normal translation, however, sometimes read through of the opal codon occurs and translation gives rise to P1234 from which the nsP4 protein gets cleaved by cis activity of nsP2 protease. Later, nsP1, nsP2 and nsP3 get separated through trans proteolysis by nsP2.

The nsP2 non-structural protein plays most important role in polyprotein processing. It cleaves P123 as well as P1234 through site specific proteolysis to produce four non-structural proteins nsP1, nsP2, nsP3 and nsP4 (Hardy & Strauss, 1989; 1988). The nsP2 acts as both cis-and trans-protease. The cis-activity occurs at early infection stage where the cleavage takes place between nsP3/nsP4 and nsP4 gets separated. Afterwards, nsP2 protein acts as trans-protease to release nsP1 from P123. Now the trans-cleavage occurs between nsP2/nsP3 by nsP2 protein and then produces nsP2 and nsP3 proteins. Thus all four non-structural proteins are produced from proteolysis of polyprotein by nsP2.

### 1.5.2 Functions of Non-Structural proteins

Non-Structural proteins function for viral replication in alphavirus life cycle. The first non-structural protein nsP1 shows guanine-7-methyltransferase and guanylyl transferase activities which are required for the capping of genome, hence considered as capping enzyme (Ahola & Kääriäinen, 1995; Mi & Stollar, 1991). It is also believed to be membrane associated and helps to

anchor replication complex to the plasma membrane (Peränen *et al.*, 1995). The nsP2 performs several functions for the viral replication as well as the non-structural polyprotein processing. It possesses helicase, NTase, RNA triphosphatase activities along with the proteolytic activity (Vasiljeva *et al.*, 2001; Vasiljeva *et al.*, 2000; Gomez *et al.*, 1999; Rikkonen *et al.*, 1994). The nsP3 has ADP-ribose 1-phosphate phosphatase and RNA-binding activity and it is also believed to be involved in virus-host interaction (Malet *et al.*, 2009; Park & Griffin, 2009). The non-structural protein nsP4 acts as RNA-dependent RNA-polymerase (RdRp) and performs very important role in virus replication (Rubach *et al.*, 2009; Tomar *et al.*, 2006; Hahn *et al.*, 1989) It is also involved in interaction with nsP1 (Shirako *et al.*, 2000). Together these imply that non-structural proteins function for alphavirus replication and are very important in virus infection (Table 1.1).

**Table 1.1:** Functions of Non-Structural proteins.

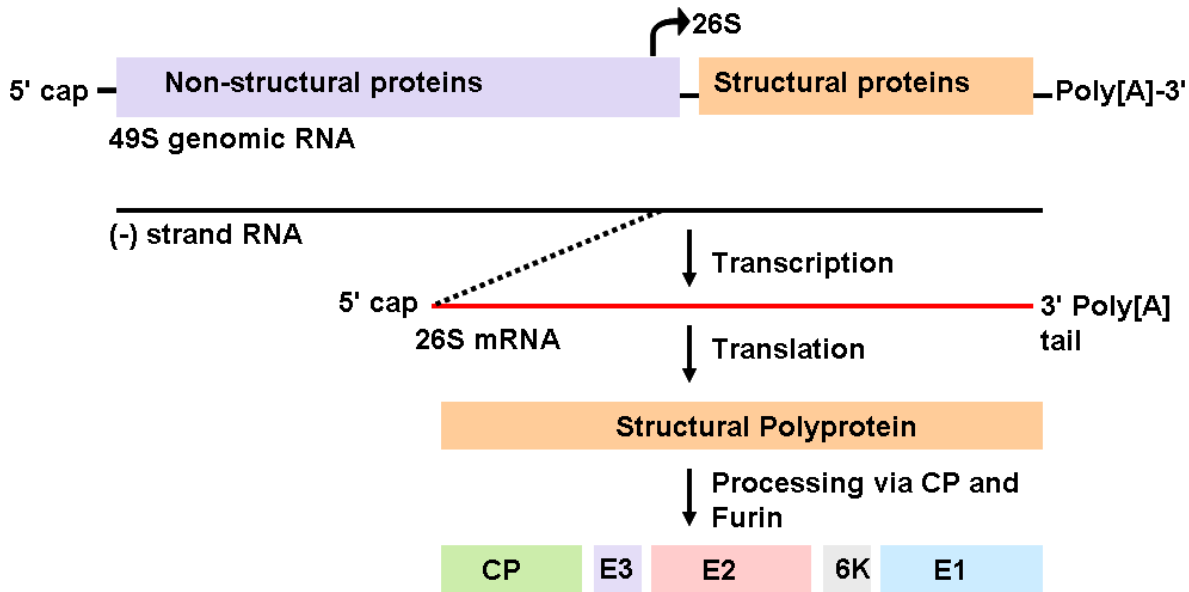
<b>Viral protein</b>	<b>Size (kDa)</b>	<b>Functions</b>	<b>References</b>
nsP1	~61	Methyltransferase and guanyltransferase, capping enzyme, Membrane association, Interaction with nsP4	Shirako <i>et al.</i> , 2000; Ahola & Kääriäinen, 1995; Peränen <i>et al.</i> , 1995; Mi & Stollar, 1991
nsP2	~90	Helicase, NTPase, nonstructural protein protease, RNA triphosphatase	Vasiljeva <i>et al.</i> , 2001; Vasiljeva <i>et al.</i> , 2000; Gomez <i>et al.</i> , 1999; Rikkonen <i>et al.</i> , 1994
nsP3	~60	ADP-ribose 1-phosphate phosphatase, RNA binding, Virus-host interaction	Malet <i>et al.</i> , 2009; Park & Griffin, 2009
nsP4	~69	RNA-dependent RNA polymerase, Interaction with nsP1	Rubach <i>et al.</i> , 2009; Tomar <i>et al.</i> , 2006; Shirako <i>et al.</i> , 2000

## 1.6 Alphavirus Structural proteins

The C-terminus one-third region of the genome encodes for the structural proteins. They form the structural part of the virus and are involved in virus budding.

### 1.6.1 Structural polyprotein processing

The genomic RNA contains the coding region for structural polyprotein. The minus strand RNA synthesis takes place by the genomic RNA which serves as the template for the transcription of 26S subgenomic RNA. The subgenomic RNA consists of both 5' cap and 3' poly-A tail and further translates into the structural polyprotein (Rice & Strauss, 1981). At the amino terminus of CP, translational enhancer is present which induces the translation by subgenomic RNA (Frolov & Schlesinger, 1994). The structural polyprotein produces different structural proteins in the order CP-E3-E2-6K-E1 through polyprotein processing by CP and furin protease.



**Figure 1.6.1.1: Processing of Structural polyprotein.** The carboxyl terminus one third region of the genomic RNA encodes for the structural polyprotein. The 49S genomic RNA produces minus strand RNA which further gets transcribed to 26S subgenomic RNA. This subgenomic RNA translates to form structural polyprotein which gets processed by CP and furin to produce the CP and the glycoproteins.

The CP, a chymotrypsin-like serine proteinase present at the N-terminus of the polyprotein, cleaves itself from the structural polyprotein precursor by cis-autoproteolysis (Tong *et al.*, 1993; Hahn & Strauss, 1990; Hahn *et al.*, 1985; Aliperti & Schlesinger, 1978). After the release of CP from the structural polyprotein, the remaining part transfers to the endoplasmic reticulum and golgi bodies co-translationally through signal peptide (Garoff *et al.*, 1978; Wirth *et al.*, 1977). After reaching endoplasmic reticulum and golgi, the furin protease (Rice & Strauss, 1981) acts on the E1-pE2 protein and cleaves it into separate E1 and pE2 glycoproteins which later form the heterodimer and act as spike proteins to envelop the virus particle (Rice & Strauss, 1982;

Ziemiecki *et al.*, 1980; Jones *et al.*, 1977). A small 6K protein also gets released from the polyprotein at the same time. In some viruses, pE2 further gets cleaved into E3 and E2 glycoproteins by furin cleavage in golgi (Jain *et al.*, 1991; Garoff & Simons, 1974). The release of E3 from pE2 precursor is highly essential as the virions containing uncleaved pE2 are non-viable (Ryman *et al.*, 2004).

### 1.6.2 Functions of structural proteins

The structural proteins function for the structure of virus and also help in virus release. The functions for the structural proteins are tabulated in Table 1.2.

**Table 1.2:** Functions of Alphavirus Structural Proteins.

<b>Viral protein</b>	<b>Size (kDa)</b>	<b>Function</b>	<b>References</b>
Capsid	~30	Formation of nucleocapsid assembly, Autoprotease activity for structural polyprotein processing, Genomic RNA encapsidation, interaction with E2 glycoprotein that leads to budding	Hong <i>et al.</i> , 2006; Choi <i>et al.</i> , 1991; Metsikkö & Garoff, 1990; Weiss <i>et al.</i> , 1989; Vaux <i>et al.</i> , 1988; Melancon & Garoff, 1987
E3	~7	Act as Signal sequence, stabilization of the heterodimer, part of E2 precursor, Interaction with E1	Uchime <i>et al.</i> , 2013; Lobigs <i>et al.</i> , 1990
E2	~47	Receptor binding and receptor mediated endocytosis, part of E1-E2 heterodimer, Interaction with CP that leads to budding	Li <i>et al.</i> , 2010; Tucker <i>et al.</i> , 1997; Lee <i>et al.</i> , 1996; von Bonsdorff & Harrison, 1978
6K	~6	Virus budding, Membrane permeabilization and virus assembly, Glycoprotein trafficking	Sanz & Carrasco, 2001; Sanz <i>et al.</i> , 1994; Lusa <i>et al.</i> , 1991; Gaedigk-Nitschko & Schlesinger, 1990
E1	~48	E1-E2 heterodimer formation, Membrane fusion protein	Li <i>et al.</i> , 2010; von Bonsdorff & Harrison, 1978

## Chapter 1: Review of literature

The CP acts as autoprotease to release itself from rest of the structural polyprotein (Choi *et al.*, 1991; Melancon & Garoff, 1987). It also encapsidates genomic RNA to form nucleocapsid assembly and interacts with the glycoproteins that leads to budding of alphaviruses (Hong *et al.*, 2006; Metsikkö & Garoff, 1990; Weiss *et al.*, 1989; Vaux *et al.*, 1988).

The cryo-EM structure of immature virus having pE2 is also available (Wu *et al.*, 2008; Ferlenghi *et al.*, 1998; Paredes *et al.*, 1998). pE2 is responsible for the proper folding of E1 glycoprotein and behaves as chaperon at the time of synthesis and transport of glycoproteins. After furin cleavage, it produces E2 and E3 glycoproteins. E3 is a very small glycoprotein and involves in the proper folding of pE2. It also interacts with E1 protein and stabilizes the heterodimer. E3 has an important role in E1-pE2 heterodimer formation and its transit towards the budding site (Lobigs *et al.*, 1990). Additionally, E3 plays an important role in pH protection during the assembly and exit of the virus (Uchime *et al.*, 2013).

E1 and E2 proteins form heterodimer and self-assemble to form trimeric spikes. These 80 trimeric spikes form the virion envelope and penetrate through the lipid bilayer (von Bonsdorff & Harrison, 1978). E2 glycoprotein helps in receptor binding and is very important for receptor mediated endocytosis. E2 is also a determinant of infection in human as well as in mosquito (Myles *et al.*, 2003; Tucker *et al.*, 1997). It is also involved in the transport of virus to the plasma membrane and virus-host interaction (Navaratnarajah & Kuhn, 2007). E2 consists of N-linked glycosylation sites which are identified through cryo-electron microscopy. These glycosylation sites play very important role in heparin sulphate binding (increase the virulence by increasing replication in host tissues) and their removal results in the increased efficiency of heparin sulphate binding (Knight *et al.*, 2009; Ryman *et al.*, 2007). Thus, E2 glycoprotein has role in neurovirulence.

6K is very small protein of 6 kDa and also present in very small amount in virion. It is essential structural component and undergoes palmitoylation which is quite important for the formation of infectious virion (Liljestrom *et al.*, 1991; Gaedigk-Nitschko & Schlesinger, 1990; Gaedigk-Nitschko *et al.*, 1990). 6K protein is involved in the transport of pE2-E1 heterodimer at the site of virus assembly (Lusa *et al.*, 1991) as can be seen from the mutational studies in 6K (Sanz & Carrasco, 2001). The 6K protein is implicated in membrane modification by ion-channel formation through lipid bilayer (Madan *et al.*, 2005; Sanz *et al.*, 2003; Melton *et al.*, 2002; Sanz *et al.*, 1994). Hence this protein can be considered as viroporins, essential for viral release, membrane permeability, glycoprotein transport and caspase-dependent apoptosis (Madan *et al.*,

2008; Gonzalez & Carrasco, 2003). During the translation of 6K protein, a frameshift event occurs which leads to the production of a novel protein named TF (transframe) which retains the ion-channel activity of 6K protein (Snyder *et al.*, 2013).

E1 glycoprotein is involved in the formation of icosahedral shell of the virus particle. It has an overall fold which is very similar to the E protein of flavivirus (Strauss & Strauss, 2001; Rey *et al.*, 1995). E1 protein is a fusion protein and responsible for the fusion of viral envelop with the endosomal membrane of host at the time of virus entry. It happens through the conversion of viral surface protein into the ion-permeable pores (Wengler *et al.*, 2003). These ion-channels make the acidic environment near endosomes which results in core disassembly. At low pH, the rearrangements in the glycoproteins take place that lead to the exposure of fusion peptide in E1 protein (Li *et al.*, 2010). However, low pH environment is not essential in mosquito cell infection (Hernandez *et al.*, 2001). Also, E1 protein consists of N-linked glycosylation sites (Pletnev *et al.*, 2001).

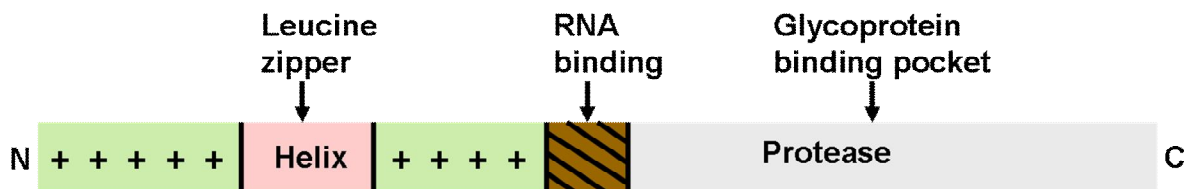
### **1.7 The nucleocapsid**

The nucleocapsid consists of genomic RNA encapsidated by the CP. The CP undergoes self assembly and form the nucleocapsid core. Like the glycoproteins, nucleocapsid core is also arranged in a T=4 icosahedral symmetry creating the projections outwards the virion, called as capsomeres.

Alphavirus CP consists of two major domains (Figure 1.7.1). The N-terminal domain is rich in basic amino acids, is highly disordered and involved in the encapsidation of RNA (Owen & Kuhn 1996; Forsell *et al.*, 1995; Geigenmüller-Gnirke *et al.*, 1993). Intrinsically disordered regions in proteins are engaged in important molecular interactions with proteins, acids or membranes that are crucial for functional activity of the protein (Uversky, 2002). The disordered N-terminal domain of the alphavirus capsid interacts with the genomic RNA and also participates in protein–protein interactions that lead to capsid dimerization. These interactions eventually result in nucleocapsid core formation (Perera *et al.*, 2001; Owen & Kuhn, 1996). Similarly, the N-terminal domain (~98 residues) of another alphavirus protein nsP4 is predicted to be disordered and also participates in protein–protein interaction (Rupp *et al.*, 2011; Tomar *et al.*, 2006). As intrinsically disordered regions perform important structural or biological functions, disordered regions in proteins including the alphavirus capsid, are therefore considered to be the potential drug targets (Metallo, 2010; Uversky *et al.*, 2008).

## Chapter 1: Review of literature

The C-terminal domain possesses the cis autoproteolytic activity and the catalytic triad residues include Ser, Asp and His which are well conserved among all the serine proteases (Choi *et al.*, 1996; Tong *et al.*, 1993; Choi *et al.*, 1991; Hahn & Strauss, 1990; Melancon & Garoff, 1987). Additionally, the CP contains a hydrophobic pocket which participates in protein–protein interaction with the glycoprotein shell that facilitates the virus budding process (Wilkinson *et al.*, 2005; Owen & Kuhn, 1997; Skoging *et al.*, 1996; Suomalainen *et al.*, 1992). In the crystal structure of Sindbis virus CP, this hydrophobic pocket was occupied by a solution-derived dioxane (Lee *et al.*, 1998). Based on this crystal structure data, dioxane-based antiviral compounds have been designed whose inhibit Sindbis virus replication (Kim *et al.*, 2007). Thus, the multifunctional CP is an excellent target for developing antiviral inhibitors to combat alphaviral diseases.



**Figure 1.7.1: Schematic representation of the CP.** The C-terminus protease domain containing glycoprotein binding site is shown in blue. The specific RNA binding region is presented in brown with slanting lines. The helix I region having leucine zipper is indicated in pink. The rest of the N-terminal region containing positively charged residues is shown in green.

The N-terminus region (residues 97-106 in SINV) is responsible for specific RNA binding (Owen & Kuhn, 1996). This region (residues 94-106 in SINV) has been suggested to act as ribosomal RNA binding domain and involved in the interaction with the ribosomes (Wengler & Wengler, 1984; Ulmanen *et al.*, 1976). This interaction has been demonstrated to be involved in disassembly, however, no role in assembly has been depicted yet (Singh & Helenius, 1992; Wengler *et al.*, 1992). A complete loss of specific RNA binding happens upon mutation in the region 76-107 of CP (Weiss *et al.*, 1994). However, the region from 1-96 amino acids in SINV is responsible for the non-specific RNA binding as this region is highly basic (Forsell *et al.*, 1995; Geigenmüller-Gnirke *et al.*, 1993; Rossmann & Johnson, 1989; Garoff *et al.*, 1980). The only conserved region among alphaviruses at the N-terminus region of CP lies between 38-55 residues (numbering according to SINV) which are present in the form of  $\alpha$ -helix named as helix I (Perera *et al.*, 2001). This region consists of the conserved Leu residues which form the leucine zipper, the mutations of those results in the reduced viral replication. The helix is suggested to be involved in inter-capsomeric interaction and assembly intermediates stabilization, thus stabilizing the virion. However, this helix can be substituted with the heterologous coiled coil and the replacement of this

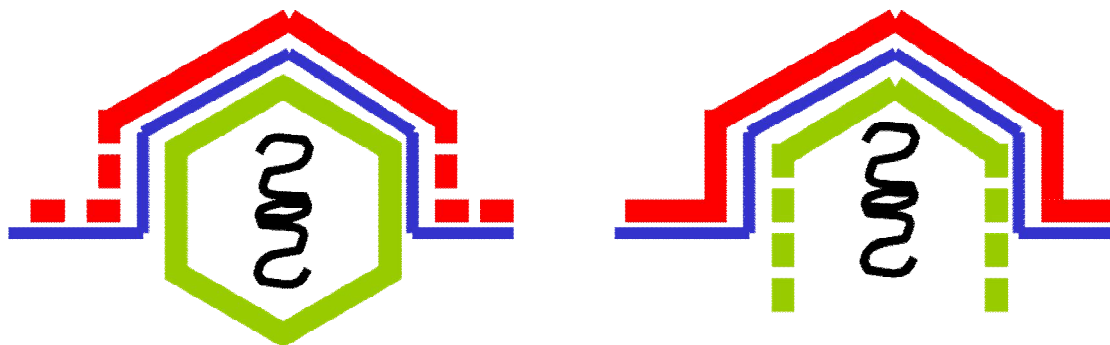


region with the unrelated sequence demonstrates the wild type phenotype with reduced viral replication (Perera *et al.*, 2003).

## 1.8 Virus budding

The two kinds of budding models have been proposed. One model suggests the formation of nucleocapsid assembly first which further promotes the budding process through binding to the glycoproteins. While the other one proposes the glycoprotein driven budding in which the glycoproteins interacts with the poorly organized capsid-RNA complex. The former one is the model described originally for the alphavirus budding (Brown, 1980; Garoff & Simons, 1974). According to this model, the nucleocapsid assembles and organizes into T=4 icosahedral symmetry which further interacts with the glycoproteins that leads to budding. This hypothesis was supported by a number of experiments. The cells infected with SFV contain a high amount of nucleocapsid-like structures in the cytoplasm and found to undergo budding at the plasma membrane (Acheson & Tamm, 1967). The non-ionic detergent extracts of the infected cells also found to contain large number of nucleocapsid (Söderlund, 1973). Nonetheless, it is very difficult to confirm the exact function of nucleocapsid as a precursor of virion.

The latter model has also been examined for confirmation. Forsell *et al.*, 2000 demonstrated the presence of budding in SFV in the absence of nucleocapsid assembly. For the experiment, a deletion mutant was prepared for SFV in which the RNA binding region and the assembly forming region was missing. The deleted region includes the RNA-binding region, the interdomain peptide as well as the leucine zipper (involved in dimerization). The amino-terminal 40 residues and the protease domain were found intact in the construct. The SFV infected BHK cells shows the presence of the viral particle with the mutant construct similar to the wild type infection. However, no nucleocapsid assembly was visible in the cytoplasm and the mutant capsid was found to be monomer. Thus, the budding process without prior nucleocapsid assembly formation is possible. The cryo-EM structure of this mutant virion has also been determined which shows the presence of glycoproteins in properly organized state with T=4 symmetry. Some other mutations were also tested and found to produce normal amount of virus but being defective due to the absence of nucleocapsid formation (Skoging-Nyberg & Liljestrom, 2001; Forsell *et al.*, 1996; Lee *et al.*, 1996). However, the exact mechanism and the driving force for budding are not clear yet.



**Figure 1.8.1: Schematic of two types of budding models.** The left panel shows the original nucleocapsid driven budding. The nucleocapsid (green) is organized first and then binds to the glycoproteins (red) 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)

### 1.8.1 Role of glycoproteins in viral budding

After structural polyprotein processing, the glycoprotein transfers towards the endoplasmic reticulum and golgi. In golgi, these undergoes glycosylation and then interacts with nucleocapsid which result in the viral budding (Strauss et al, 1995; Zhao & Garoff, 1992). The post-translational modification occurs during the translocation of the spike proteins. This includes the palmitoylation of the cysteine residues 396, 416 and 417 present at the cytoplasmic tail of E2 glycoprotein in SINV (Ivanova & Schlesinger, 1993; Gaedigk-Nitschko & Schlesinger, 1991; Liljeström & Garoff, 1991). The interaction of capsid and glycoprotein occurs through the cytoplasmic tail of the E2. The mutagenesis studies demonstrate the role of the cytoplasmic tail residue Tyr400 of E2 in SINV and SFV. For proper virus propagation, this residue is essential and can be replaced only by the hydrophobic residue. The studies suggest that this residue plays important role for CP-glycoprotein interaction (Zhao *et al.*, 1994; Gaedigk-Nitschko & Schlesinger, 1991). In one other study, it has been demonstrated that the phosphorylation and dephosphorylation of this residue in E2 cytoplasmic tail occurs which is essential for virus budding. The kinase inhibitors are found to show the defect in viral budding. These inhibit the interaction of glycoproteins with the capsid at the plasma membrane (Liu *et al.*, 1996; Liu & Brown, 1993). Other mutational studies suggest that the conserved residues Tyr-X-Leu are important for binding to the hydrophobic pocket of the capsid (Skoging-Nyberg & Liljeström, 2000; Owen & Kuhn, 1997).

Moreover, it was previously assumed that the cytoplasmic domain of E1 has no role in CP-glycoprotein interactions (Barth *et al.*, 1992). However, Tang *et al.* (2011) demonstrated that the interaction of cytoplasmic tail of E1 and CP might occur and they 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). An anti-idiotypic antibody was developed and found to interact with capsid, this revealed the interaction of E2 cytoplasmic tail and capsid (Vaux *et al.*, 1988). However, in later studies, this antibody was found to interact with capsid non-specifically and directed towards the RNA replication site (Suomalainen & Garoff, 1992).

A high resolution cryo-EM structure of Venezuelan equine encephalitis virus (VEEV) has been solved at 4.4 Å (Zhang *et al.*, 2011). The cryo-EM structure shows the endo as well as ectodomains of the E1 and E2 glycoproteins. In this report, the cytoplasmic domains of both the glycoproteins E1 and E2 have been modeled. The structure reveals several interesting features regarding capsid and the glycoproteins which were previously unknown. The E3 protein which was found to be cleaved in mature VEEV previously has been identified in mature VEEV cryo-EM. Furthermore, one additional helix was observed in CP which is found in the interdomain region of the capsid and might be involved in interaction with the host ribosome during disassembly.

### **1.8.2 Role of 6K in virus budding**

The 6K peptide interferes with the CP-glycoprotein interaction through alteration in the ionic condition at the site of budding. The 6K action as the ion channel has also been shown in the artificial membrane (Melton *et al.*, 2002). The 6K peptide is not the part of the release virion, however, it translocates with the glycoproteins to the plasma membrane (Lusa *et al.*, 1991; Gaedigk-Nitschko & Schlesinger, 1990). Thus, the 6K peptide has been described as to play an important role in virus budding (Ivanova *et al.*, 1999; Yao *et al.*, 1996; Loewy *et al.*, 1995; Liljestrom *et al.*, 1991).

### **1.8.3 Role of virus RNA in budding**

The genomic RNA encapsidation occurs specifically through interaction between the non-structural region of genomic RNA and the RNA binding domain of the CP (Frolova *et al.*, 1997; Weiss *et al.*, 1994; Weiss *et al.*, 1989). In the absence of this specific interaction in defective virions, the subgenomic RNA can also be encapsidated by CP. However, Aura virus CP naturally encapsidates subgenomic 26S RNA (Rumenapf *et al.*, 1995). The nucleocapsid formation requires the capsid-RNA interaction as the *in vitro* assembly of the nucleocapsid was observed after the

## Chapter 1: Review of literature

interaction of CP with RNA either specifically or non-specifically (Tellinghuisen *et al.*, 1999; Wengler *et al.*, 1982). As the RNA encapsidation directs nucleocapsid assembly, hence it is indirectly involved in the viral budding. However, budding also can occur without the formation of nucleocapsid assembly according to the model described above (Forsell *et al.*, 2000).

### 1.9 Structure of alphavirus capsid protein

The crystal structures of the CP from different alphaviruses are available. The details are provided in Table 1.3. All the crystal structures of CP from alphaviruses either obtained from the wild type protein (extracted from virus) or the recombinant protein, are very similar to each other and do not show any structural variations (Figure 1.9.1). The crystal structure of alphavirus CP consists of two  $\beta$ -barrel domains arranged in a Greek key motif. At the interface of the two domains, catalytic triad is present. The catalytic triad residues include Ser, His and Asp which are present far apart in the sequence, however, present very near in the structure to form the active site. After the autoproteolysis, the C-terminus Trp remains bound near the active site and thus makes the protein further inactive. Therefore, the protein is active just for one reaction in the life cycle. SCP has the shorter loops between the  $\beta$ -strands as compared to the chymotrypsin.

The CP encapsidates the genomic RNA through its highly basic N-terminal domain and forms the nucleocapsid which further gets self-assembled to form the nucleocapsid assembly. This N-terminal region was found to be disordered in different crystal forms of alphaviruses. The crystal structure of the full length CP is also available, but it reveals only the electron density for the C-terminus protease domain only. It is not confirmed whether the N-terminal region is disordered or deleted during the process of crystallization. On SDS-PAGE a very faint band was there, due to which it was very difficult to predict the molecular weight of the protein from the full length crystal (Choi *et al.*, 1996).

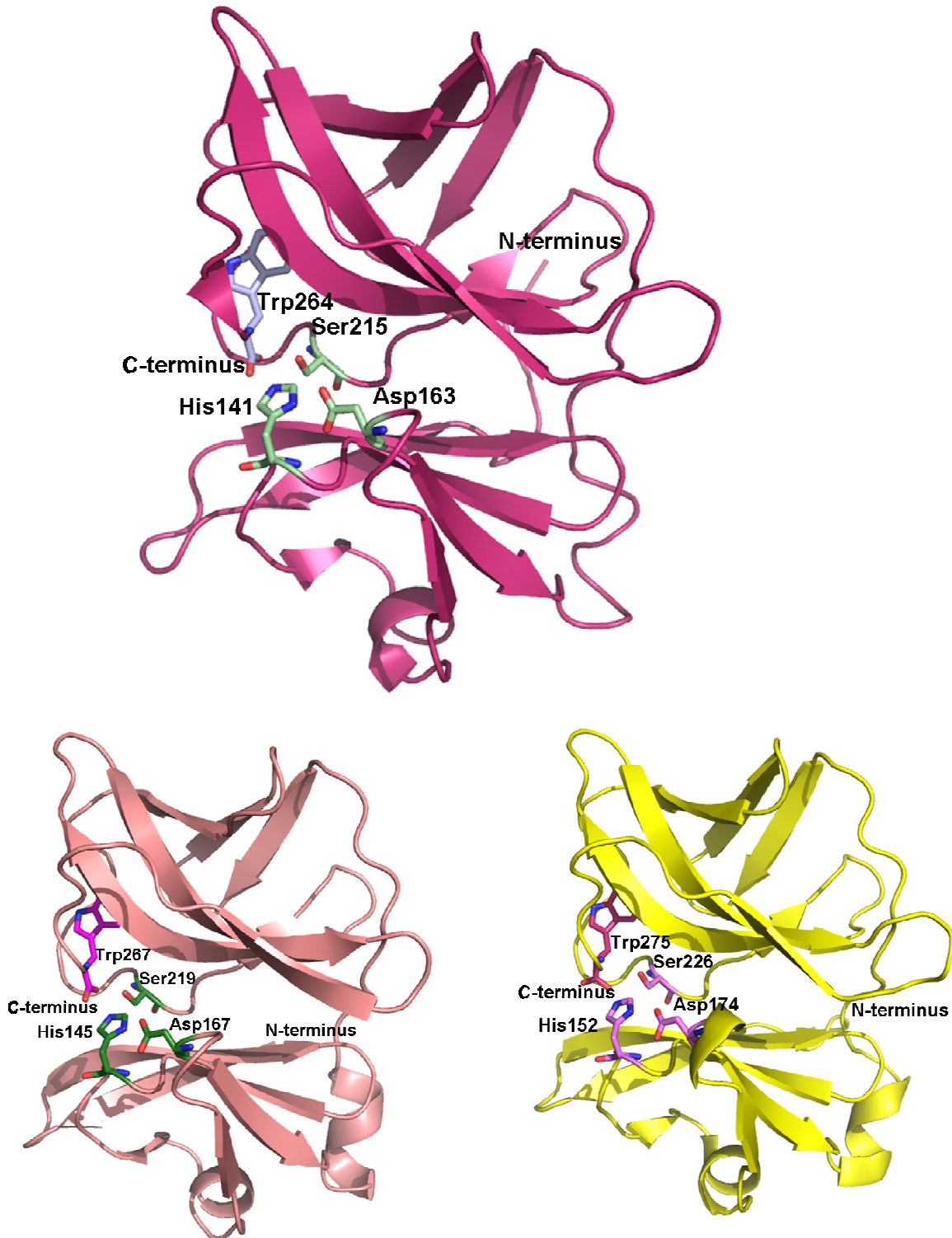
The CP exists as monomer, dimer, trimer as well as tetramer in different crystal forms. However, the dimer is absent in the viral core assembly and the solution form (Morillas *et al.*, 2008; Cheng *et al.*, 1995). The monomeric units of CP from all alphaviruses have similar overall fold. Sindbis virus CP (SCP), Semliki Forest virus CP (SFCP) and VEEVCP crystal forms show monomer, dimer, trimer as well as tetramers (Table 1.3). The dimer interface area in SCP is 380  $\text{\AA}^2$  which is very small in comparison to the other oligomeric proteins (Tong *et al.*, 1993; Choi *et al.*, 1991). The interface residues that are involved in dimer formation include residues 185 to 190, 194 to 195 and 222 in SCP. The interaction includes both H-bonding as well as hydrophobic

interactions. The hydrogen bond forming residues Asn190 and Asn222 were mutated and the structure of mutants does not show any change in the oligomeric state. However, the mutation in the hydrophobic residue Phe188 leads to crystallographic dimer disruption. Though, any of these mutants do not show any reduction in viral replication or assembly (Choi *et al.*, 1996).

**Table 1.3:** CP structural data from different members of alphavirus family.

Virus	PDB ID	Features	Year	Resolution (Å)	Number of monomers per asymmetric unit
SINV	2SNV		1993	2.8	1
	1KXE	(Y180S, E183G)	1996	3.2	1
	1KXD	(N222L)	1996	3.0	1
	1KXC	(N190K)	1996	3.1	1
	1KXB	(S215A)	1996	2.9	1
	1KXA		1996	3.1	1
	1KXF	(1-264)	1996	2.38	1
	2SNW		1998	2.7	2
	1WYK	(dioxane bound)	1998	2.0	4
	1SVP	(substrate bound)	1996	2.0	2
SFV	1VCP		1996	3.0	3
	1VCQ		1996	3.1	2
VEEV	1EP5		2003	2.3	3
	1EP6		2003	2.45	3

The cytoplasmic domain of the CP is also involved in the interaction with E2 glycoprotein that results in viral budding. A hydrophobic pocket is present which is surrounded by the hydrophobic residues. E2 protein interacts with CP via this hydrophobic pocket. In the crystallographic symmetry molecules, the hydrophobic pocket of one molecule interacts with the amino-terminus leucine residues of other neighboring molecule which mimics the CP-glycoprotein interaction (Lee *et al.*, 1996).



**Figure 1.9.1. The crystal structures of the CP from different alphaviruses.** The SCP (PDB ID: 1KXA) is shown as dark pink cartoon form, SFCP (PDB ID: 1VCP) in light pink and VEEVCP (PDB ID: 1EP5) in yellow. The overall fold of all the crystal structures is similar and contains two  $\beta$ -barrel subdomains at the interface of which the catalytic triad is present. The catalytic triad residues are shown as sticks. The C-terminal region turns towards the active site for approaching the substrate to the active site as well as specificity pocket. After autoproteolysis, the C-terminus Trp remains bound near the active site and is static in conformation in all the crystal forms (Choi *et al.*, 1997; 1991).

### 1.10 Role of Capsid protein in virus budding

Alphaviruses are icosahedral enveloped viruses of approximately 700 Å in diameter (Jose *et al.*, 2009). The nucleocapsid core is surrounded by the phospholipid bilayer, in which trimeric spikes of E1 and E2 heterodimers are embedded (Mukhopadhyay *et al.*, 2002; Cheng *et al.*, 1995). The interaction of the E2 glycoprotein with the capsid takes place through the binding of E2 tail in the hydrophobic pocket of the CP which results in budding (Sjöberg and Garoff, 2003; Cheng *et al.*, 1995; Strauss *et al.*, 1995; Lee *et al.*, 1994; Lopez *et al.*, 1994; Vaux *et al.*, 1988). Several studies have been done to find out the interacting region in the E2 glycoprotein and the studies suggest the presence of conserved Tyr-X-Leu motif in the E2 cytoplasmic tail. This conserved motif has been found to be involved in the interaction with the CP hydrophobic pocket (Skoging-Nyberg & Liljeström, 2000; Levine *et al.*, 1996; Zhao *et al.*, 1994; Kail *et al.*, 1991; Metsikkö & Garoff, 1990).

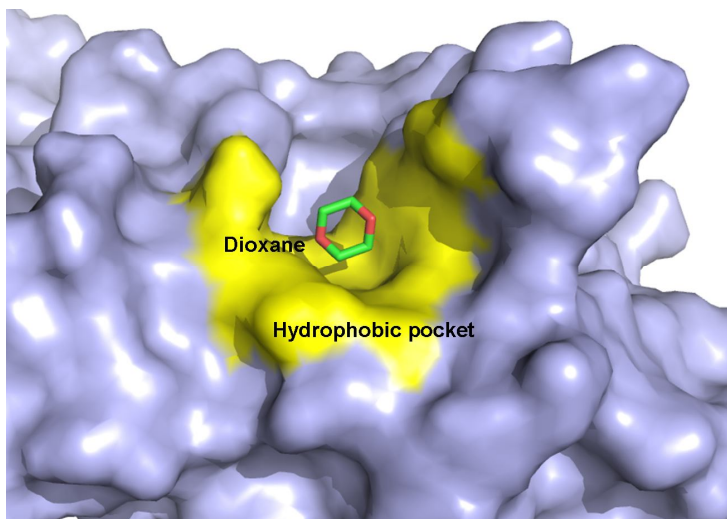
As the name suggests, the hydrophobic pocket of the CP consists of the hydrophobic residues and creates the hydrophobic environment for the proper interaction with the E2 glycoprotein cytoplasmic tail. One study suggests that the E1 glycoprotein might also be involved in the interaction with the CP during budding and plays significant role for the budding of virus (Tang *et al.*, 2011). In this report, several mutations have been performed to get the knowledge of the residues important in the interaction with E2 glycoprotein and involve in virus budding. Furthermore, the cryo-EM structure of VEEV also provides the information regarding the interaction of the CP and cytoplasmic domain of the glycoproteins (PDB ID: 3J0C). The structure shows the clear interaction of the E2 glycoprotein in the hydrophobic pocket of the CP.

The analysis of the nucleocapsid residues involve in the interaction with glycoprotein demonstrates the presence of both T=4 and T=7 virions upon mutations of residues Tyr180 and Glu183 (Y180S and E183G) (Lee & Brown, 1994). These mutations make the virion temperature sensitive and at the temperature different from normal, the non-infectious virions are generated. This might be due to the defective disassembly process. The studies on this double mutant virion also showed the difference in the cross-linking between the capsid and glycoprotein than the wild type (Lee & Brown, 1994). Skoging *et al.*, 1996 also confirmed the role of Tyr180 in capsid-glycoprotein interaction. Upon deletion of Tyr180, the empty pocket has been occupied by Trp247 along with the structural variations. The binding of the N-terminal arm to the hydrophobic pocket of the neighboring molecule is not influenced by the mutations in the double mutant (Choi *et al.*, 1996). The second mutation E183G disrupts the salt bridge with Arg225, both these residues are

## Chapter 1: Review of literature

involved in monomer-monomer contact in nucleocapsid (Cheng *et al.*, 1995) and hence, this second mutation leads to the formation of T=7 particles along with T=4 virions.

Additionally, the crystal structure of CP from Sindbis virus was found to contain the solvent-derived dioxane in its hydrophobic pocket (Lee *et al.*, 1998) (Figure 1.10.1). This suggested that dioxane or similar molecules may be able to enter the pocket and prevent capsid-E2 binding. In order to target and disrupt CP-glycoprotein interactions, dioxane based synthetic antiviral compounds have been designed against Sindbis virus based on the crystal structure of SCP containing dioxane in the hydrophobic pocket (Kim *et al.*, 2007; 2005).



**Figure 1.10.1:** Hydrophobic pocket of SINV CP with bound dioxane (PDB ID: 1WYK) (Lee *et al.*, 1998)

The cryo-EM structure of the SINV has been determined at a resolution of 7 Å. In this structure, the interaction of the transmembrane E1 and E2 glycoproteins with the CP has been demonstrated and the mutational studies revealed several interactions and regions which are essential for CP-glycoprotein interaction and for budding (Tang *et al.*, 2011). The amino-terminus region of the CPs forms an adjoining network for the connection of CP pentamers and hexamers. The study depicts the interaction of a single glycoprotein to three adjacent CP capsomeres which can be considered as the process for budding initiation. In SINV, the residues Arg393 and Glu395 of E2 and Tyr162 and Lys252 of CP were found to be essential for virus budding.

### 1.11 RNA encapsidation and capsid assembly

The N-terminal motif (Lys97-Met106) of SCP was found to be responsible for specific encapsidation as this motif has been shown to be required for specificity of genomic RNA encapsidation (Owen & Kuhn, 1996). The amino-terminus residues before this region are not

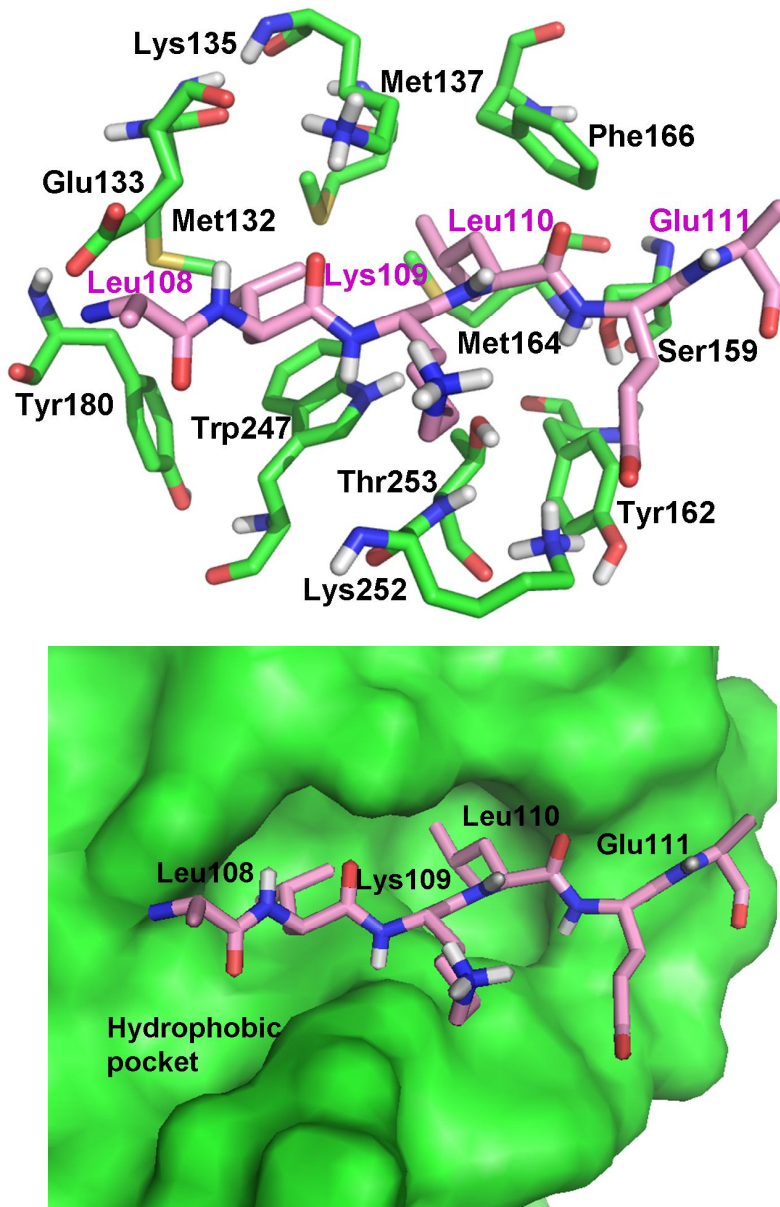


conserved among the alphaviruses and responsible for the non-specific interaction with the RNA. The structures of the DNA binding domain of proteins from bacteria have been determined in complex with the DNA (Narayanan *et al.*, 2014; 2012). However, this is pretty difficult with less stable RNA.

After autoproteolysis reaction, CP is found to be associated with ribosomes (Söderlund & Ulmanen, 1977; Glanville & Ulmanen, 1976; Söderlund, 1973). However, the exact role of ribosomal association in nucleocapsid assembly is poorly understood. Also after entry, the nucleocapsid core remains bind to the ribosome. This interaction was believed to play important role in virus disassembly (Singh & Helenius, 1992; Wengler *et al.*, 1992).

The crystal structure of SCP indicates the interaction of 108-110 amino terminus residues to the hydrophobic pocket of the neighboring molecule (Figure 1.11.1). This hydrophobic pocket is same in which the cytoplasmic tail of E2 glycoprotein binds. Thus, it has been suggested that this interaction between the N-terminus residues of the CP with the hydrophobic pocket mimics the binding of E2 glycoprotein to the same hydrophobic pocket. The Tyr180 residue in SCP is buried in the crystal structure. Moreover, the mutational studies for these N-terminal residues support their role in capsid assembly (Lee *et al.*, 1996). However, this interaction with the hydrophobic pocket was found to be absent in SFV (Choi *et al.*, 1997). This was considered as the reason for the exposure of the corresponding residue Tyr184 in the absence of bound glycoproteins in SFCP. However, this might be due to different crystal packing in SFCP.

The residues Leu108 and 110 mimic the E2 residues Tyr400 and Leu402 and might be involved in the similar interactions with the hydrophobic pocket. The mutational studies of the E2 glycoproteins show the presence of similar infectious virions if these residues have been mutated to hydrophobic residues (Zhao *et al.*, 1994). Thus only hydrophobic residues can fit into the hydrophobic pocket of the CP and leucine fits the best in this category. Hence, it has been hypothesized that the amino terminal arm of the CP itself binds to the hydrophobic pocket in the same manner as E2 glycoprotein binds which was further confirmed by cryo-EM structure of the alphaviruses (Cheng *et al.*, 1995).



**Figure 1.11.1:** The N-terminal arm of the CP binds to the hydrophobic pocket of the neighboring CP which mimics the interaction of E2 glycoprotein with CP. The N-terminal arm of one molecule of the CP is shown as pink sticks. The hydrophobic residues are presented as sticks (upper panel) and spheres (lower panel) representing the hydrophobic pocket (green in color). This interaction is supposed to play an important role in virus assembly. (Lee *et al.*, 1996)

## 1.12 Viral proteases

Chymotrypsin like serine proteases have been reported in vertebrates, bacteria as well as in viruses. The crystal structures of proteases from different sources are known and are almost similar to each other having chymotrypsin like fold. The alphavirus capsid protease and the hepatitis C virus NS3 protease come in this category of viral proteases (Choi *et al.*, 1997; 1996; Kim *et al.*, 1996). All the structures consist of two Greek key  $\beta$ -barrel subdomains. The active site consists of

the catalytic triad comprising Ser, His and Asp residues. Serine acts as the catalytic nucleophile in the active site. The active site is located at the interface of two subdomains. The catalytic triad residues are located far away from each other in the sequence, but due to the folding of the structure these residues come near to each other. In fact, their relative orientation in all the serine proteases remains the same. For viral proteases, induced fit mechanism has been demonstrated (Li & Jing, 2000; Landro *et al.*, 1997; Kim *et al.*, 1996).

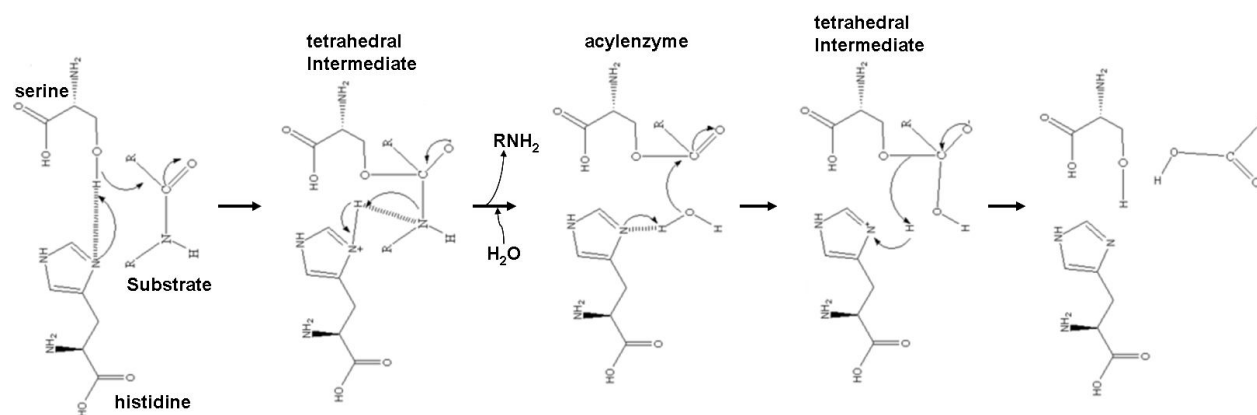
### **Mechanism of the action of Serine proteases**

The catalytic triad residue aspartate forms H-bond with histidine through its carboxyl group and thus makes nitrogen atom of histidine more electronegative. When the substrate binds to the serine protease, its scissile bond comes in contact with the active site and brings its carbonyl carbon atom towards the nucleophilic serine. After the entry of the substrate -OH group of serine attacks on the carbonyl carbon atom of the substrate and its hydrogen is accepted by the nitrogen atom of histidine which is electronegative due to aspartate. In the scissile bond, the electron pair from the double bond of carbonyl group moves towards oxygen. All these give rise to a tetrahedral intermediate stage formation. The covalent bond electrons of the substrate scissile bond move towards the hydrogen of histidine as a result of which the break down of the peptide bond of the substrate occurs. And the electrons from the carbonyl oxygen move back to form the double bond again. This is considered as the acyl enzyme intermediate (Figure 1.12.1).

As this is a hydrolytic reaction, now water comes to the reaction which replaces the N-terminus part of the peptide and the same reaction proceeds again as before; the movement of the electrons towards carbonyl oxygen, acceptance of proton from water molecule by histidine and the generation of tetrahedral intermediate. In the last step of the reaction, the movement of electron occurs from the bond between substrate carbon and oxygen atom of serine to hydrogen atom of the histidine making the carbonyl carbon of substrate more electron deficient. It induces the formation of double bond again and now the other half of the peptide also gets separated from the active site serine.

The oxyanion hole helps in stabilizing the tetrahedral intermediate stages. This consists of Gly and Ser which form H-bonds by donating their backbone hydrogen atoms to the intermediates. Both the intermediates have the negatively charged carbonyl oxygen which fits perfectly into the oxyanion hole. This transition state of the serine protease is highly preferable due to low activation energy.

## Chapter 1: Review of literature



**Figure 1.12.1:** Catalytic mechanism of the serine proteases. Two tetrahedral intermediates are formed and the product is released in two steps. (Hedstrom, 2002)

Other proteases derived from animals have been studied in detail for the characterization of protease activity (Kumar *et al.*, 2007; 2006; Mohanty *et al.*, 2003). The parasitic cysteine proteases have also been isolated and characterized (Yadav *et al.*, 2007). Protease inhibitors are attractive target for designing and development of new therapeutic agents (McKinlay, 2001; Matthews *et al.*, 1999; Wang, 1999; Patick *et al.*, 1998). Several inhibitors for different serine proteases are known which can be either chemical inhibitors or the peptide inhibitors. These can resemble the tetrahedral stage also which will compete with the substrate for binding to the active site. The structure of the HCV NS3 protease in complex with its inhibitor has been solved by nuclear magnetic resonance spectroscopy and X-ray crystallography (Barbato *et al.*, 2000; di Marco *et al.*, 2000). Similar to the chymotrypsin, in NS3-inhibitor complex, the inhibitor is found to be present in extended conformation. The FRET (fluorescence resonance energy transfer) technique has also been used for inhibitor binding studies (Fattori *et al.*, 2000).

### 1.12.1 Capsid protein as serine protease

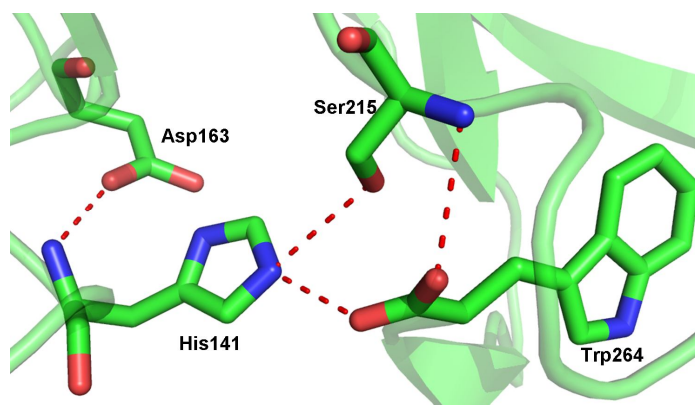
The C-terminal domain of CP possesses the cis autoproteolytic activity and becomes inactive after self cleavage due to the binding of carboxyl terminus tryptophan residue to the active site (Skoging & Liljestrom, 1998; Choi *et al.*, 1996; Tong *et al.*, 1993; Choi *et al.*, 1991). The active site conformation and the catalytic triad residues Ser, Asp and His are well conserved among all the serine proteases. The scissile bond is present at the C-terminal end where CP cleaves between WS residues that are conserved among all alphaviruses. This tryptophan residue is crucial for autoproteolytic activity. On mutation of the tryptophan to Ala, Phe or Arg, there is complete

loss of the protease activity. The viral growth is efficient only in the wild type form having tryptophan at the C-terminus (Skoging & Liljestrom, 1998). The Trp to Ala mutation makes the protease completely inactive (Thomas *et al.*, 2010). The sequence Gly-Asp-Ser-Gly is conserved among all chymotrypsin like serine proteases which consists of catalytic serine residue.

### 1.12.1.1 The catalytic triad of alphavirus CP

The catalytic triad of the alphavirus CP consists of Ser, His and Asp like other serine proteases. The mutation of the serine residue results in complete loss of protease activity (Melancon & Garoff, 1987). Ile227 plays an important role in the protease activity of Chikungunya virus CP. The mutation I227K through error prone PCR leads to the inactivation of the protease activity of the enzyme, because of its position near the active site and Lys was suggested to disrupt the electrostatic interaction of the catalytic triad residues (Thomas *et al.*, 2010). The alphavirus CP functions as chymotrypsin like serine protease. Other chymotrypsin like proteases, hepatitis A virus 3C (HAV3C) and Rhinovirus 3C (HRV3C) contains cysteine residue in place of serine at the active site. However, the overall fold is quite similar to SCP (Allaire *et al.*, 1994; Matthews *et al.*, 1994).

The C-terminus tryptophan remains bound near the active site in all the crystal forms as the crystal structures available are for the cleaved protein. The active site Ser215 shows interaction with this tryptophan as shown in Figure 1.12.1.1.1. The interaction takes place due to the binding of scissile bond. After cleavage tryptophan remains bound and does not get separated due to which the protein is further inactive. The three active site residues also show interactions with each other and remain at the same position in all the crystal forms.



**Figure 1.12.1.1.1: Active site of SCP.** The catalytic triad residues are presented as sticks. The interaction between the catalytic triad residues His141, Asp163 and Ser215 is shown in red dashed lines. Trp264 also reaches towards the active site and forms polar contacts with the catalytic triad residues His and Ser. (Tong *et al.*, 1993)

## Chapter 1: Review of literature

The structural arrangement for the catalytic triad might vary between different proteases, while they possess the very similar relative arrangement. The crystal structure shows almost similar distance between the catalytic triad residues as present in chymotrypsin. However, other differences and deletions exist. Ser214 residue in chymotrypsin interacts with the catalytic triad residue Asp102, however the corresponding residue Leu231 in SCP does not show such interaction with Asp163. Asp163 is partially exposed to the solvent in SCP (Tong *et al.*, 1993).

### 1.12.1.2 Oxyanion hole

The oxyanion hole consists of residues Gly213 and Gly216 in SCP which are conserved among all the serine proteases. In HAV3C, the mutation in the catalytic triad residue cysteine shows conformational change in the oxyanion hole (Allaire *et al.*, 1994). However, the mutant SCP (S215A) does not depict any conformational variation as compared to the native form (Choi *et al.*, 1996). Subtilisin also showed similar results when the equivalent active site serine was mutated (Carter & Wells, 1988).

The substrate bound SCP crystal structure has also been solved. In this crystal form, the SCP construct consists of two extra residues at the C-terminus and S215A mutation was done at the active site. Due to this mutation, the protein is not able to show autoproteolytic activity and the last two residues are not get cleaved. Because this crystal form contains the residues that act as substrate for the protein, this form has been described as substrate bound form. This form of SCP is similar to the other wild-type forms. However, the two chains of this form show structural variations at the C-terminus. The oxyanion hole contains a water molecule in chain A which is similar to the structure of chymotrypsin with no bound substrate. The water molecule is absent from chain B where the oxyanion hole residues form H-bonds with the O atom of the main chain preceding scissile bond (Choi *et al.*, 1996). Thus, chain A has been referred as early stage of catalysis with loosely bound substrate.

### 1.12.1.3 Specificity pockets of alphavirus CP

Different specificity pockets can be seen in the crystal structure of alphavirus CP. The specificity pockets are the substrate binding pockets to which the incoming substrate binds. The residues involved in forming these substrate binding pockets (specificity pockets) of SCP are tabulated in Table 1.4.

**Table 1.4:** Residues involved in the formation of different specificity pockets.

Specificity pocket	SCP Residues involved
S <sub>4</sub>	Ala234, Ser243, Pro258
S <sub>2</sub>	His141
S <sub>1</sub>	Trp191, 205-215, 229-235, 240-242
S' <sub>1</sub>	Ile126, His141, Val142
S' <sub>2</sub>	Val125, Ile126
S' <sub>4</sub>	Asn120, Asp124

The S<sub>1</sub> specificity pocket includes the residues which are involved in binding to the tryptophan. In SCP this pocket includes Trp191, 205-215, 229-235, 240-242 residues. The pocket consists of the crucial residue which is important for the substrate specificity. Trypsin contains Asp189 in the S<sub>1</sub> pocket which is essential for binding to the lysine or arginine group of substrate and thus maintains substrate specificity. However, the chymotrypsin like serine proteases consist of small hydrophobic amino acid, so that the ample space can be provided to the tryptophan residue of the substrate. The corresponding residue is serine in chymotrypsin and valine in SCP.

The S<sub>1</sub> specificity pocket is hydrophobic in SCP and plays the most important role in catalysis as involved in binding to the P<sub>1</sub> residue tryptophan. It protects tryptophan from the solvent and maintains Trp at its position. In all the crystal forms of SCP, the position of Trp264 is static and does not show any conformational variations. However, the two chains of the substrate bound mutant form of the capsid demonstrate large conformational change of the substrate bound to the specificity pocket. In chain A, the substrate was found to be far away from the active site and oxyanion hole; while in chain B, it is present near the active site important for catalysis. The two chains depict a large deviation in the substrate residues. That's why the chain A has been described as the early stage of catalysis and the substrate is loosely bound to it.

Thr261-Glu-Glu-Trp264 forms the structure above the surface of the CP similar to the substrate in other proteases. As the Trp264 occupies the S<sub>1</sub> pocket, likewise Glu263, Glu262 and Thr261 reside into the S<sub>2</sub>, S<sub>3</sub> and S<sub>4</sub> pocket respectively. The side chains of Glu263 and His141 are parallel to each other while Glu262 is directed outwards towards the solvent. The S<sub>4</sub> pocket consists of residues Ala234, Ser243 and Pro258. The autoproteolytic cleavage occurs between Trp264 and Ser265 (E3 residue). After P'<sub>1</sub> residue Ser265; Ala266, Ala267 and Pro268 are present which are considered as P'<sub>2</sub>, P'<sub>3</sub> and P'<sub>4</sub> residues. These residues occupy different specificity pocket

## Chapter 1: Review of literature

as determined by comparing the structure from other chymotrypsin like serine proteases. S'<sub>1</sub> consists of residues Ile126, His141 and Val142. The S'<sub>2</sub> pocket contains Val125 and Ile126, while S'<sub>4</sub> pocket has Asn120 and Asp124 residues (Tong *et al.*, 1993).

### 1.12.1.4 Trans-activity of capsid protein

The CP from alphaviruses shows cis-autoproteolytic activity. An attempt has made to observe its trans-catalytic activity, however, the trans activity was not observed and the protein was considered as the enzyme which is active for just one reaction in the virus life cycle (Hahn & Strauss, 1990). Almost two decades later, further studies were performed to find out the possibilities of the trans activity of the alphavirus CP. In this view, the truncations were performed in the capsid which includes deletion of last Trp residue along with other C-terminus residues (Morillas *et al.*, 2008). After these deletions, the activity assays for the trans-activity for the protein have been performed and the protein was found to be fully active with these truncations. The esterase activity of the capsid protease was shown to be very high, however, the trans-proteolytic activity has not been observed. After performing a number of experiments viz. circular dichroism (CD), fluorescence and Nuclear magnetic resonance (NMR), the truncated construct of the protein was depicted as the natively unfolded form present in the inclusion bodies. This indicates that the C-terminus Trp residue plays a most important role in proper folding of the CP structure (Morillas *et al.*, 2008). As this protein was found in the inclusion bodies, so it was very difficult to solubilize and purify the active protein. To overcome the problem, the denaturants were used for expression and purification of the protein. Also, the  $K_{cat}/K_M$  value of the enzyme was found to be  $5 \times 10^5 \text{ s}^{-1}\text{M}^{-1}$  which is very similar to the values for other serine proteases (Powers & Kam, 1995; Odake *et al.*, 1991; Lottenberg *et al.*, 1981).

The NMR, fluorescence and CD experiment of the protein provide the details of the disorderedness of CP after C-terminal truncations including last Trp residue. The C-terminally deleted constructs have the same structural features as that of the denatured native protein. Thus, it has been suggested that the active protein with deleted residues at the C-terminus, is natively unfolded. However, it is highly active and shows efficient catalysis of tyrosine or tryptophan nitrophenyl ester substrates. Previously, a number of cases have been observed which shows the unfolded or very less folded protein (Dyson & Wright, 2005; Fink, 2005; Uversky *et al.*, 2000). Even existing as natively unfolded form, the proteins undergo their functions properly and thus the



functional property of these proteins were termed as their intrinsic feature (Liang *et al.*, 2007; Dafforn & Smith, 2004; Uversky, 2002).

Several FRET based proteolytic assays have been observed till date, which use the fluorogenic peptide substrate that contains a FRET pair on both the sides (Mittoo *et al.*, 2003; Angliker *et al.*, 1995; Knight *et al.*, 1992). After protease action, the substrate gets cleaved and the FRET is not observed. The protease enzyme activity of SARS (Severe acute respiratory syndrome) coronavirus 3CL protease has been determined using FRET based substrate. This protease has a catalytic dyad made up of Cys and His, the mutations in any of the residue results in the inhibition of proteolytic activity. The fluorogenic substrate was having the peptide that contains the scissile bond and is specific for this protease. This technique can be used further for the testing of compounds against the protease activity of the enzyme (Chen *et al.*, 2005). A high-throughput drug screening technique can be developed by using these FRET based peptide substrates for antiviral drug discovery.

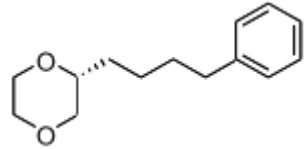
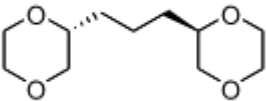
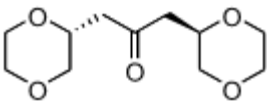
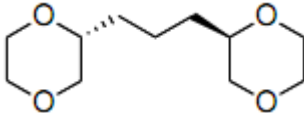
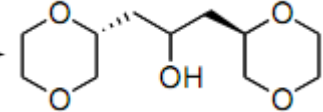
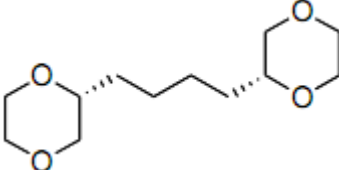
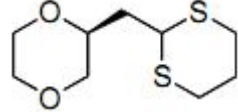
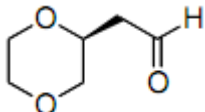
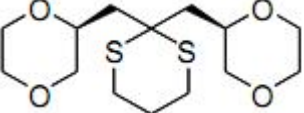
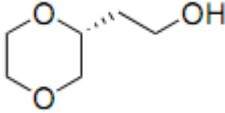
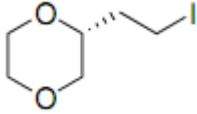
### **1.13 Strategies to inhibit alphavirus capsid protein**

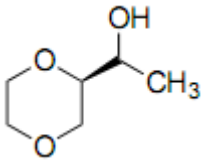
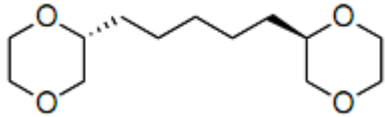
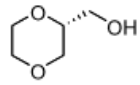
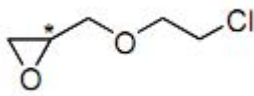
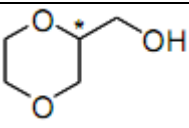
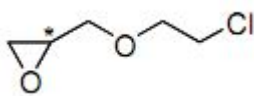
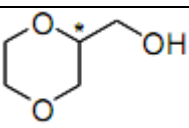
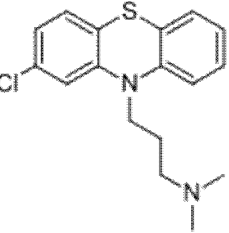
Two different inhibition strategies were tested to find out the antiviral drugs against the alphavirus CP: the one includes the inhibition of the CP-glycoprotein interaction through competitive inhibition for the hydrophobic pocket binding and the other one targets the protease activity of the CP. Some successful attempts were done for the hydrophobic pocket binding, however, the protease inhibition efforts could not achieve any antiviral activity.

#### **1.13.1 Protein-protein interaction inhibition through dioxane analogs**

The alphavirus CP consists of the hydrophobic pocket to which the glycoprotein binds to the capsid. Interestingly, a solvent-derived dioxane molecule was found to be bound to this hydrophobic pocket in the crystal structure of SCP (Lee *et al.*, 1998), suggesting that dioxane derivatives may prevent capsid-E2 hydrophobic interactions (Figure 1.10.2). As expected, dioxane-derived molecules showed antiviral properties against Sindbis virus (Kim *et al.*, 2007; 2005). Therefore, dioxane and its derivative molecules are expected to bind the hydrophobic pocket and inhibit virus budding by disrupting capsid-E2 interaction in alphaviruses. After testing the dioxane derivatives as antiviral agents, a number of these compounds were found to be potent antivirals. These compounds and their antiviral activity are given in Table 1.5.

**Table 1.5:** Dioxane analogs and their antiviral activity (Kim *et al.*, 2007; 2005).

S. No.	Compound Structure	EC50 (M)	CC50 (M)
1		$>1 \times 10^{-4}$	$0.6 \times 10^{-3}$
2		$4 \times 10^{-5}$	$>1 \times 10^{-3}$
3		$>1 \times 10^{-4}$	$>1 \times 10^{-3}$
8		$1.4 \times 10^{-5}$	$>1 \times 10^{-3}$
10		$6.7 \times 10^{-4}$	$>1 \times 10^{-3}$
16		$1.5 \times 10^{-3}$	$>1 \times 10^{-3}$
20		$1 \times 10^{-3}$	$>1 \times 10^{-3}$
21		$4.6 \times 10^{-4}$	$>5 \times 10^{-3}$
22		$2.9 \times 10^{-5}$	$>1 \times 10^{-3}$
27		$1.4 \times 10^{-3}$	$>1 \times 10^{-3}$
29		$2.5 \times 10^{-4}$	$>1 \times 10^{-3}$

33		$9.9 \times 10^{-5}$	$\sim 1 \times 10^{-3}$
36		$1.5 \times 10^{-3}$	$> 1 \times 10^{-3}$
(R)-7		$1 \times 10^{-6}$	$> 1 \times 10^{-3}$
(R)-3		$1 \times 10^{-3}$	$> 1 \times 10^{-3}$
(R)-4		$3.4 \times 10^{-6}$	$> 1 \times 10^{-3}$
(S)-3		$1 \times 10^{-3}$	$> 5 \times 10^{-3}$
(S)-4		$1.5 \times 10^{-3}$	$> 5 \times 10^{-3}$
Chlorpromazine		$8.2 \times 10^{-6}$	$1 \times 10^{-4}$

### 1.13.2 Capsid protease inhibition through the substrate-based inhibitors

The substrate based inhibitors can bind to the specificity pocket to block the entry of the substrate molecule. The lack of the crystal structure of the active form of CP hampers the design of structure based inhibitors. Different inhibitors based on the esterase activity of the CP were tested for the inhibition of enzymatic activity but no inhibition has been observed. Some trypsin inhibitors are known that have been isolated from plant sources and characterized for the inhibitory activity against trypsin (Patil *et al.*, 2012; Chaudhary *et al.*, 2008). Chloromethylketones like N-benzyloxycarbonyl-leucine-tyrosine-chloromethylketone and N-benzyloxycarbonylphenylalanine-chloromethylketone and Phenylmethylsulphonyl fluoride (PMSF) have been considered as potent

## Chapter 1: Review of literature

inhibitors of serine proteases (Johnson & Moore, 2002; Johnson & Moore, 2000; Jung *et al.*, 1995; Eguchi & Kuriyama, 1985; Sidorowicz *et al.*, 1980). However, these inhibitors do not affect the enzymatic esterase activity of SFCP (Morillas *et al.*, 2008). Also, the serine protease inhibitor AEBSF (4- (2-Aminoethyl) benzenesulfonyl fluoride hydrochloride) did not show the inhibition of CHIKV capsid protease activity (Thomas *et al.*, 2010).

## **2.1 Abstract**

This chapter represents the crystal structure of Aura virus capsid protein (AVCP) and the homology models of E1 and E2 glycoproteins. The structure of AVCP is compared to the capsid protein (CP) structures from other alphavirus members. The AVCP crystal structure and the homology models of E1 and E2 glycoproteins have been fitted into the cryo-electron microscopic (cryo-EM) density map of Venezuelan equine encephalitis virus (VEEV) to find out the interaction between the capsid and glycoproteins. The sequence analysis has also been performed for the studies of different regions of the protein and for the comparative analysis between different members of the alphavirus genus. The structural analysis reveals new insights into the interaction between the capsid and glycoproteins, important for virus budding.

The C-terminal protease domain of AVCP has been expressed in a bacterial expression system and purified to homogeneity. The purified protein was crystallized. Crystals suitable for X-ray diffraction analysis were obtained by the vapor-diffusion method using 0.1 M Bis-Tris and Polyethylene glycol monomethyl ether 2000. The crystal belonged to the space group *C2*, with unit-cell parameters  $a = 79.6$ ,  $b = 35.2$ ,  $c = 49.5$  Å. High-resolution data sets were collected to a resolution of 1.81 Å. Preliminary crystallographic studies suggested the presence of a single molecule in the crystallographic asymmetric unit, with a solvent content of 38.5 %.

The nucleocapsid core interaction with endodomains of glycoproteins plays a critical role in the alphavirus life cycle that is essential to virus budding. Recent cryo-electron microscopy studies provide structural insights into the key interactions between the CP and the trans-membrane glycoproteins E1 and E2. The CP possesses a chymotrypsin-like fold with a hydrophobic pocket at the surface responsible for interaction with glycoproteins. In the absence of crystal structures, homology models of E1 and E2 from Aura virus were generated. The crystal structure of CP and structural models of E1 and E2 were fitted into the cryo-EM density map of VEEV for detailed analysis of CP-glycoprotein interactions. Structural analysis reveals that the E2 endodomain consists of a helix-loop-helix motif where the loop region fits into the hydrophobic pocket of CP. Our studies suggest that Cys397, Cys418 and Tyr401 residues of E2 are involved in stabilizing the structure of E2 endodomain. Density map fitting analysis reveals that Pro405, a conserved E2 residue is present in the loop region of the E2 endodomain helix-loop-helix structure and makes intermolecular hydrophobic contacts with the capsid.

### 2.2 Introduction

Alphaviruses are the causative agents of diseases ranging from mild fever to harsh encephalitis and may also be a menace of bioterrorism (Reichert *et al.*, 2009; Sherman & Griffin, 1990). The Chikungunya virus epidemics and the increasing number of cases of Chikungunya infection day by day create the threat to human health (Simon *et al.*, 2011; Hoarau *et al.*, 2010; Schwartz & Matthew, 2010). Hence, there is an urgent need for the development of effective antiviral therapies and drugs against alphaviruses. Also, complete understanding of the alphavirus budding process will provide valuable information, as the budding process is a potential drug target.

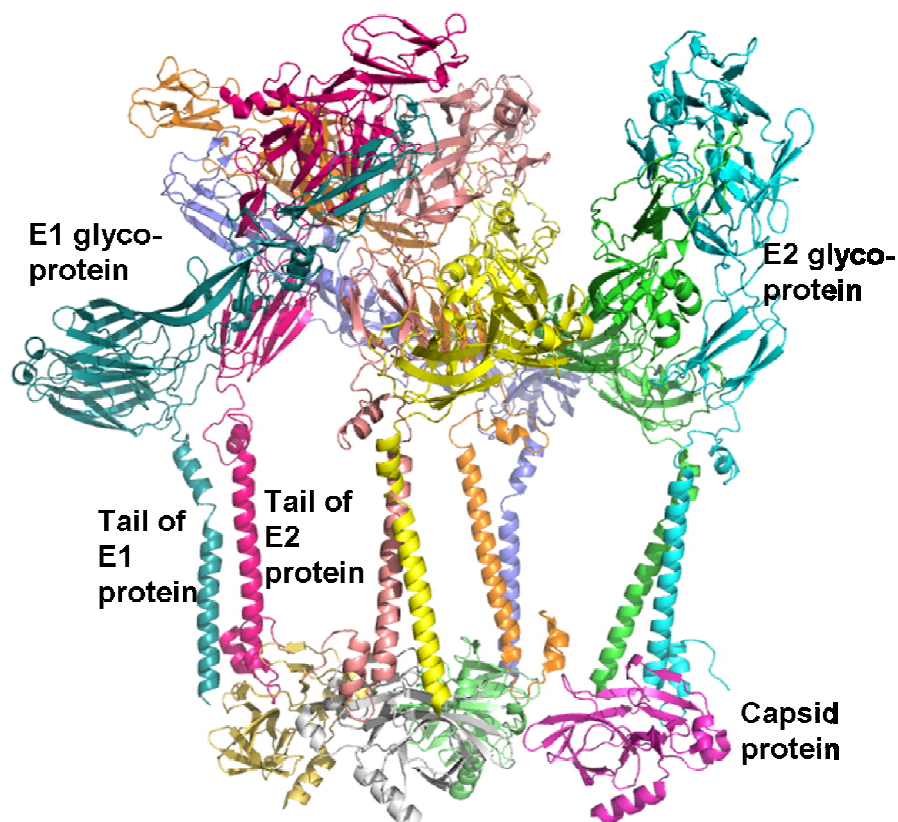
The alphaviruses contain a nucleocapsid core surrounded by the lipid envelope through which spike glycoproteins penetrate. The nucleocapsid core is formed by the encapsidation of RNA by CP. The E2 glycoprotein interacts with the nucleocapsid complex via a hydrophobic pocket present in the carboxyl-terminal region of the CP that leads to budding of alphaviruses (Sjöberg & Garoff, 2003; Cheng *et al.*, 1995; Strauss *et al.*, 1995; Lee *et al.*, 1994; Lopez *et al.*, 1994; Vaux *et al.*, 1988). Earlier investigations have shown that different regions in the cytoplasmic tail of E2 (cdE2) are involved in CP-glycoprotein interaction (Levine *et al.*, 1996; Kail *et al.*, 1991; Metsikkö & Garoff, 1990). In this view, the conserved Tyr-X-Leu motif in E2 glycoprotein has been suggested to play a direct role in the interaction with the hydrophobic pocket of CP (Skoging-Nyberg & Liljeström, 2000; Zhao *et al.*, 1994).

In the recent years, extensive studies on the CP-glycoprotein interaction in alphaviruses have riveted attention to further investigate and formulate new antiviral molecules (Jose *et al.*, 2012; Tang *et al.*, 2011; Zhang *et al.*, 2011). In fact, in these studies, the pseudo-atomic model of Sindbis virus and E2 mutational studies revealed loop regions and other conserved residues those are essential for the interaction. Furthermore, three different contact regions in the CP were identified, two of which consist of the exposed loops at the surface (Tang *et al.*, 2011). However, in order to evaluate the potency of CP-glycoprotein interaction, a more comprehensive study at the molecular level is essential.

Recently, the cryo-EM structure of (VEEV) has been determined at 4.4 Å resolution and reveals the arrangement of trans-membrane helices and cytoplasmic tails of E1 and E2 glycoproteins (Zhang *et al.*, 2011) (Figure 2.2.1). This is the highest resolution cryo-EM structure among alphaviruses structures known so far. Also, the cytoplasmic tails of the glycoproteins visible in this cryo-EM structure were not modeled in any of the cryo-EM structure for other

alphaviruses previously. According to this cryo-EM structure, the linker region (residues 115–124) of CP was in the form of an  $\alpha$ -helix and overlaps the region (residues 109–125) expected to interact with the 60S ribosomal subunit of host cell during disassembly of nucleocapsid. Interestingly, this linker region is not found to be in helical form in any known crystal structures of alphavirus CP.

This structure shows the E3 protein in mature virion which was previously considered to be deleted from mature virus particle. The E3 glycoprotein gets cleaved by furin and separated from p62 proteins. In the VEEV cryo-EM structure, the cleaved E3 has been found which remains associated with E2 protein. The biochemical analysis also reveals the presence of the cleaved E3 protein.



**Figure 2.2.1:** Cartoon view of the cryo-EM structure of VEEV (PDB ID: 3J0C). Different chains are shown in different colors. The capsid protein is present at the bottom. The cytoplasmic tails of E1 and E2 glycoproteins appears to come in contact with the hydrophobic pocket of capsid protein. The E1-E2 heterodimer is clearly visible (Zhang *et al.*, 2011).

Moreover, it was previously assumed that the cytoplasmic domain of E1 has no role in CP-glycoprotein interactions (Barth *et al.*, 1992). However, recent findings provide evidence for the interaction between the cytoplasmic tail of E1 and CP, suggesting that E1 may also have a role in

## **Chapter 2: Crystal structure of Aura virus capsid protein and insight into the capsid-glycoprotein interaction**

the viral budding process (Tang *et al.*, 2011). Although the direct evidence suggesting the role of E1 and CP interaction in virus budding is not available and needs more detailed studies.

Furthermore, a charged interaction has also been reported to occur between the amino terminus of the cytoplasmic tail of E2 and CP (Jose *et al.*, 2012). Nonetheless, both the molecular interaction of E2 with the nucleocapsid core and the role of E1 glycoprotein in budding through interaction with CP are still not clearly understood. Additional structural investigations of CP-glycoprotein interactions from different alphavirus members are required for the detailed understanding of the budding process, which could be beneficial in the development and design of new drug molecules against alphavirus infection.

The cysteine residues present at the cytoplasmic domain of the E2 glycoprotein undergoes palmitoylation. This creates the hydrophobic environment which helps in binding to the hydrophobic pocket of the CP. Thus, the process of palmitoylation helps in CP-glycoprotein interaction which further leads to budding of the virus particle (Ivanova & Schlesinger, 1993). Furthermore, In SINV E2 glycoprotein cytoplasmic tail residue Pro404 was found to play an important role in the virus infection. The growth of virus is slow in the mutant P404A and also these mutants are thermolabile in comparison to the wild type virus (Ryan *et al.*, 1998; Ivanova & Schlesinger, 1993).

### **2.3 Materials and Methods**

#### **2.3.1 Materials**

All the enzymes for cloning were procured from NEB (New England Biolabs). The primers were purchased from Integrated DNA Technologies. For protein purification, Ni-NTA beads were purchased from Qiagen and imidazole (low absorbance at 280 nm) was obtained from Acros. ÄKTA Prime plus system and HiLoad 16/60 Superdex 75 gel filtration column were acquired from GE Healthcare. Amicon ultra protein concentrators were purchased from Millipore (Bedford, Massachusetts, USA). All other chemicals were of analytical grade and purchased from commercial sources. The crystallization trays and screens were purchased from Hampton Research.

#### **2.3.2 Strategy used for the study**

The strategy used to perform the studies has been presented in Figure 2.3.2.1.



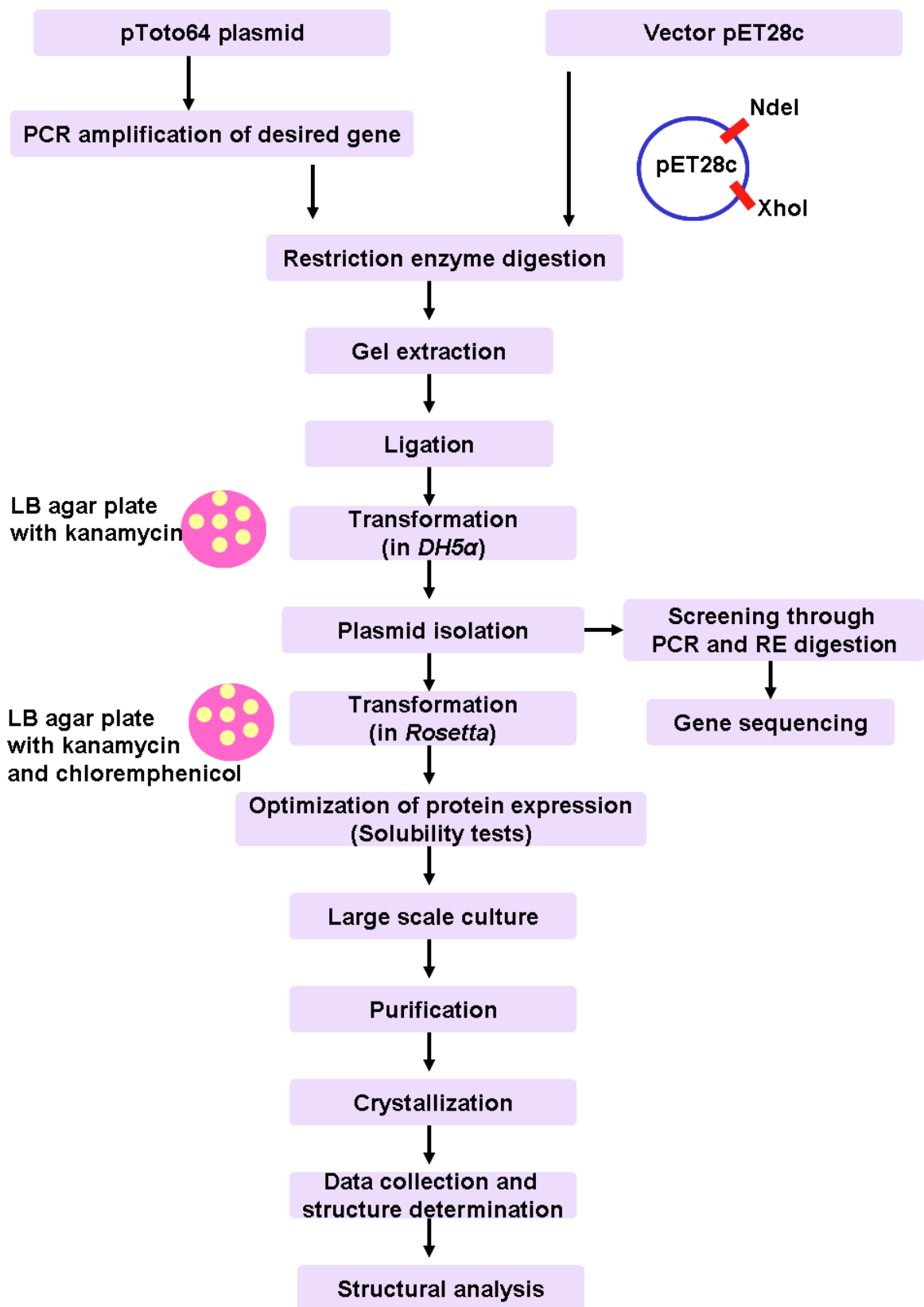


Figure 2.3.2.1: Schematic representation of the methodology used for the study of AVCP.

## Chapter 2: Crystal structure of Aura virus capsid protein and insight into the capsid-glycoprotein interaction

First of all the desired gene encoding AVCP has been cloned into the expression vector. The over-expression of the protein was done in *E. coli* strain *Rosetta* and was optimized for the better solubility of the protein. The protein was further purified using column chromatography techniques. The purified protein was then concentrated and crystallized. The diffraction data were collected and processed. After integration and scaling, the structure was determined by molecular replacement method. The refinement of the structure was carried out and the structural analysis has been performed. For the glycoproteins, the homology models were generated and both the structures were fitted into the cryo-EM density map of VEEV for the structure based study of CP-glycoprotein interaction. The fitted structure was further analyzed for the CP-glycoprotein interaction.

### 2.3.2.1 Construction of expression plasmids

Aura virus genomic cDNA (Rümenapf *et al.*, 1994) was used as the template for polymerase chain reaction (PCR) amplification of the DNA fragments encoding the full-length CP and the carboxy-terminal capsid protease domain (residues 110–267). The oligonucleotide primers 5' -ACGAACATATGAACTCTGTCTTTTACAATCCGTT- 3' (forward), 5' -AAGCACTCGAGT TACCACTCTACAGTATCTTCGTGGG- 3' (reverse) and 5' -CTGGAATTCATATGGCCCTG AAATTTGAAGCCGAC- 3' (forward), 5' -CTAGATACTCGAGCTACCACTCTACAGTATCT TCGTGG- 3' (reverse) containing NdeI and XhoI sites were synthesized for the amplification of DNA fragments encoding the full length and the capsid protease domain, respectively. The primers were designed on the basis of GenBank accession No. NP\_819015. The NdeI and XhoI restriction sites introduced into the primers allowed the cloning of these DNA fragments into the corresponding sites of the pET28c vector in which a thrombin protease cleavage site had been substituted by a TEV (tobacco etch virus) protease cleavage site.

The resulting recombinant proteins had an additional 21 amino-acid residues at the N-terminus including the His6 tag and TEV protease cleavage site (MGSSHHHHHSSENLYFQGHM). The DNA fragments were amplified by PCR using these primers and purified using a PCR purification kit (Qiagen, USA) according to the manufacturer's instructions to remove unincorporated nucleotides and primers. The purified PCR fragment and pET28c vector were restriction-digested with NdeI and XhoI restriction enzymes. The digested DNA fragments were separated by 1 % agarose gel electrophoresis, purified using a DNA gel extraction kit (Qiagen, USA) and ligated with T4 DNA ligase. *E. coli DH5 $\alpha$*  competent cells were

## **Chapter 2: Crystal structure of Aura virus capsid protein and insight into the capsid-glycoprotein interaction**

transformed by the heat-shock method (Inoue *et al.*, 1990) using the ligation mixture. Some of the colonies obtained by plating the transformed cells on Luria–Bertani (LB) agar plates containing 50 µg/ml kanamycin were picked and grown overnight. Plasmids were isolated using a MiniPrep plasmid-isolation kit (Qiagen, USA) and screened for the presence of the constructs by PCR and restriction-enzyme digestion. The integrity of the resulting plasmids was confirmed by sequencing in both directions using T7 promoter and T7 terminator primers.

### **2.3.2.2 Expression of the capsid protein**

For the production of N-terminally His-tagged recombinant AVCP, *Rosetta* (DE3) cells were transformed with plasmid containing the AVCP-encoding DNA fragment. Bacterial cultures of the transformed cells were grown in LB broth supplemented with 50 µg/ml kanamycin and 35 µg/ml chloramphenicol at 37 °C to an optical density of 0.6 at 600 nm (OD<sub>600</sub>). Expression was then induced with 0.4 mM isopropyl β-d-1-thiogalactopyranoside (IPTG) and the culture was grown for 4 hrs at 37 °C after induction. Finally, the cells were harvested by centrifugation at 6000 rpm and 4 °C and the pellets were stored at -80 °C until further use.

### **2.3.2.3 Purification of AVCP**

The cell pellet from 1 liter culture was resuspended in binding buffer (50 mM Tris–HCl pH 7.6, 15 mM imidazole, 100 mM NaCl) and lysed using a cell disruptor (Constant Systems Ltd, Daventry, England). The cell lysate was centrifuged at 14,000 rpm for 45 min at 4 °C. The clarified supernatant was loaded and incubated for half an hour on a gravity-flow column containing Ni–NTA beads (Qiagen, USA) pre-equilibrated with binding buffer at 4 °C. The recombinant CP was eluted with 250 mM imidazole in 50 mM Tris–HCl pH 7.6, 100 mM NaCl. The fractions containing purified protein were pooled and dialyzed overnight at 4 °C against dialysis buffer consisting of 50 mM Tris–HCl pH 7.6, 20 mM NaCl, 3 mM EDTA. The TEV protease was added to the protein sample in a 1:25 ratio for His-tag cleavage. The EDTA was removed from the protein sample by dialysis in the same buffer but without EDTA. A reverse Ni–NTA column was run to remove the uncleaved protein, cleaved His tag and His-tagged TEV protease. His-tag-cleaved AVCP present in the flow-through was collected and concentrated to ~ 6 mg/ml. The protein was further purified by gel-filtration chromatography using a pre-equilibrated HiLoad Superdex 75 16/60 column (GE Healthcare) and ÄKTA purifier system (GE Healthcare) that was operated at 4 °C with a flow rate of 0.5 ml/min. The size-exclusion column was calibrated

## **Chapter 2: Crystal structure of Aura virus capsid protein and insight into the capsid-glycoprotein interaction**

with an LMW Calibration Kit containing bovine serum albumin (66 kDa), ovalbumin (45 kDa), trypsin (23 kDa) and lysozyme (14 kDa) for determination of the void volume, construction of the standard curve and estimation of the molecular weight of the purified protein. The gel-filtration eluate was collected in 2 ml fractions and the purity of the fractions was analyzed by Coomassie Blue stained SDS-PAGE. The fractions containing pure protein sample were then pooled and concentrated to ~ 10 mg/ml using an Amicon Ultra-15 concentrator with a cutoff value of 3 kDa (Millipore, Bedford, Massachusetts, USA). The concentration of the purified protein was determined by UV absorbance spectroscopy at 280 nm using a calculated extinction coefficient of  $22,460 \text{ M}^{-1} \text{ cm}^{-1}$ .

### **2.3.2.4 Crystallization**

The purified protein was crystallized using the sitting-drop vapour-diffusion method. The protein concentration for crystallization was ~ 10 mg/ml in 50 mM Tris-HCl pH 7.6, 20 mM NaCl. Crystal screens from Hampton Research were used for optimization of the crystal-growth conditions. The protein and the reservoir buffer were used in a 1:1 ratio and were equilibrated against 50  $\mu\text{l}$  reservoir buffer. Crystals were grown in 100 mM Bis-Tris pH 6.5, 28 % (w/v) Polyethylene glycol monomethyl ether 2000 at 20 °C.

### **2.3.2.5 Data collection**

For high-resolution data collection, the composition of the cryo-protectant was optimized by testing various cryo-protection agents. The AVCP crystals were first briefly soaked in 15 % (w/v) glycerol for 1–2 s; this was followed by crystal annealing and the crystal was then re-soaked in 12.5 % (w/v) ethylene glycol prior to X-ray data collection. X-ray data were collected at the home source with a MAR 345 imaging-plate system using Cu K $\alpha$  radiation generated by a Bruker-Nonius Microstar H rotating-anode generator operated at 45 kV and 60 mA. The data were collected under cryogenic conditions (100 K) at a wavelength of 1.5418 Å. The diffraction data were processed with the HKL-2000 package (Otwinowski & Minor, 1997).

### **2.3.2.6 Structure determination and refinement**

The structure was solved by the molecular replacement method using the Molrep program of the CCP4 program suite (Vagin & Teplyakov, 1997; Collaborative Computational Project Number 4 1994). The crystal structure of Sindbis virus capsid protein (SCP) (PDB ID: 1KXA)

## **Chapter 2: Crystal structure of Aura virus capsid protein and insight into the capsid-glycoprotein interaction**

(Choi *et al.*, 1996) was used as a search model for structure solution. Further refinement was carried out by the REFMAC program (Murshudov *et al.*, 1997) of the CCP4 suite and the program COOT (Emsley & Cowtan, 2004). Six groups of TLS were chosen according to TLSMD server and used for anisotropic refinement of the molecule (Winn *et al.*, 2000). Iterative cycles of the refinement were performed and interspersed with visual inspection and manual adjustments to obtain acceptable values of Rcryst and Rfree. PROCHECK program was used for the evaluation of the stereo-chemical properties of the model (Laskowski *et al.*, 1993). Visualization of the refined model and model building were carried out in COOT and the figures were prepared using PyMol (Delano, 2002).

### **2.3.2.7 Molecular modeling**

Molecular modeling of E1 and E2 was carried out by comparative methods and energy minimization using the program Swiss-Model in automated mode and MODELLER 9v8 (Schwede *et al.*, 2003; Sali & Blundell, 1993). Following the five sequential steps (template selection from RCSB PDB database, sequence alignment, model generation, refinement and validation), three dimensional homology models of E1 and E2 of Aura virus were generated. The template was chosen on the basis of sequence identity and structural completeness, thus the complete VEEV structure (PDB ID: 3JOC) was used as a template for E1 and E2 model generation. MULTALIN server was used for the query sequence alignment to the template sequence (Corpet, 1988). After some manual corrections in the sequence alignment, five models were generated separately for both proteins using MODELLER. The quality of the models was assessed by PROCHECK. The best model was selected and energy minimization was performed using Swiss-PdbViewer 4.01 (<http://spdbv.vital-it.ch/>). Model validation was performed using ProSA energy plot and VERIFY-3D of the SAVES server (Wiederstein & Sippl, 2007; Luthy *et al.*, 1992). Finally, the crystal structure of AVCP and homology models of E1 and E2 were fitted into the cryo-EM density map of VEEV (EMDB ID: 5275) using fit-in-map module of Chimera software (Pettersen *et al.*, 2004). The figures of aligned sequences were generated using ClustalW and ESPript (Gouet *et al.*, 1999; Thompson *et al.*, 1994).

## **2.4 Results and Discussion**

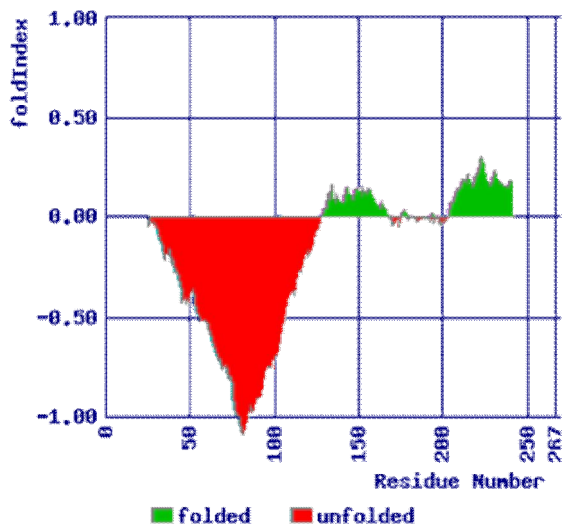
The amino-terminal region of alphavirus CP is involved in the interaction with genomic RNA, while the C-terminal region acts as a serine protease having a chymotrypsin-like fold.

## Chapter 2: Crystal structure of Aura virus capsid protein and insight into the capsid-glycoprotein interaction

Earlier reports suggested the amino-terminal region of this protein to be highly disordered based on the absence of its electron density in the crystal structure of full length CPs (SCP and Semliki Forest Virus CP (SFV)) (Choi *et al.*, 1997; Choi *et al.*, 1991). Additionally, mass spectroscopic analysis of the SFCP crystals revealed that the molecular weight of protein molecules in the crystal is the same as that of the C-terminal protease domain (Choi *et al.*, 1997).

### 2.4.1 Prediction of the disordered region in AVCP

The amino-terminal domain of SCP and SFCP is highly unstable and gets degraded during the process of purification or crystallization. The disordered region of AVCP was predicted using FoldIndex server (Prilusky *et al.*, 2005). According to the FoldIndex results, a highly disordered region is present at the amino-terminus of the protein (Figure 2.4.1.1). The intrinsically disordered region of the protein might cause difficulty in getting the over-expressed soluble protein and can cause trouble in purification and crystallization. However, their deletion from the rest of the protein might overcome these problems. Considering these facts, we expressed, purified and crystallized the C-terminal protease domain (residues 110–267) of AVCP for our studies in which the amino terminus disordered region has been truncated.

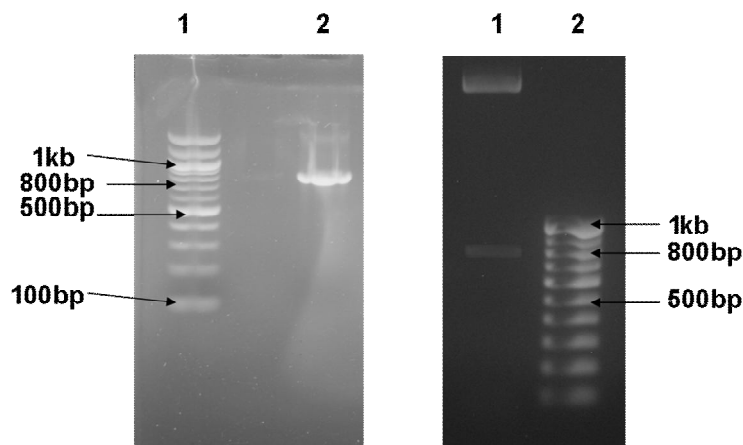


**Figure 2.4.1.1:** Structurally disordered region prediction in AVCP. The red color shows the disordered region while the green region is ordered. These predictions were done using FoldIndex server.

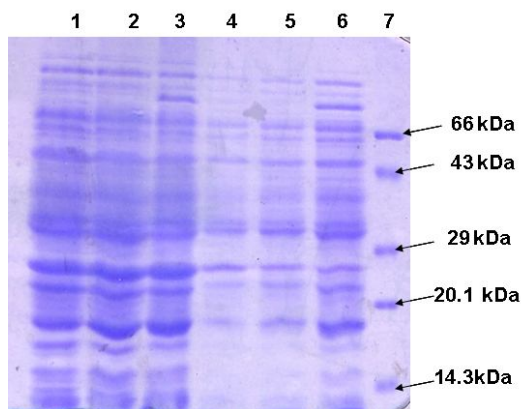
### 2.4.2 Cloning and heterologous expression of the capsid protein

The full-length AVCP-encoding gene was cloned into pET28c vector for expression in bacterial system. The PCR amplification using the primers specific for full length AVCP and the restriction enzyme digestion of cloned plasmid is shown in Figure 2.4.2.1. However, the full-

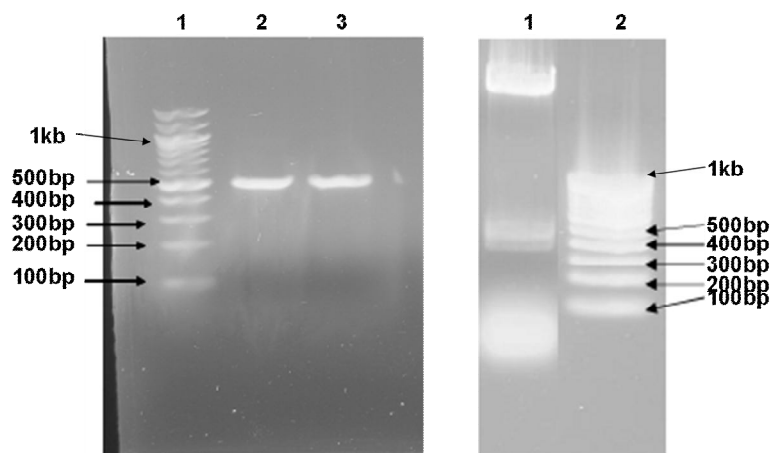
length recombinant CP did not show any expression (Figure 2.4.2.2); this may be the consequence of the presence of an unstructured amino-terminal RNA-binding domain in AVCP. Therefore, the N-terminal domain (residues 1–109) of AVCP was deleted. Another construct containing the carboxy-terminal protease domain (residues 110–267) was designed and cloned into pET28c vector for bacterial expression. The PCR amplification of the protease domain was performed using the already cloned cDNA construct of Aura virus (Figure 2.4.2.3). The cloned plasmid was screened for the confirmation of positive clones (Figure 2.4.2.3). The recombinant plasmids were sequenced and the sequencing results showed that the Val120, Gly121, Val190, Val212, Val240, Pro241 and Gly242 residues of the GenBank sequence were substituted by Ala120, Val121, Gly190, Ala212, Gly240, Ala241 and Arg242 in the AVCP construct. AVCP showed the over-expression in *E. coli* cells *Rosetta* with a very good amount of soluble protein (Figure 2.4.2.4).



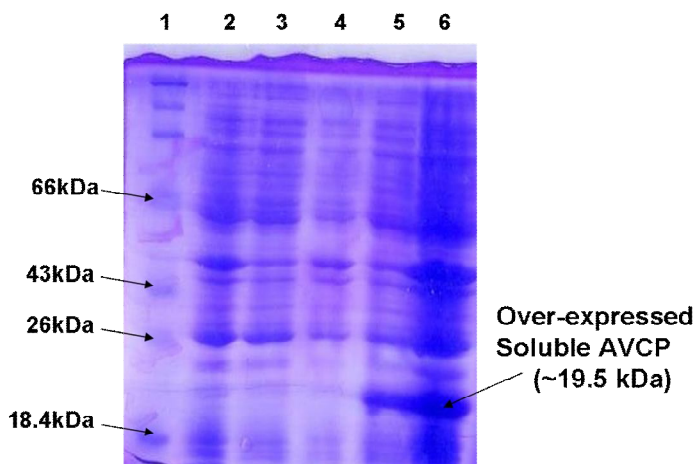
**Figure 2.4.2.1:** The PCR amplification of the full length AVCP (left panel). Lane 1 shows the DNA ladder and lane 2 represents the PCR amplified product. The right panel shows the restriction enzyme digestion of the cloned plasmid confirming the presence of correct clone.



**Figure 2.4.2.2:** Expression of the full length AVCP. The IPTG induced cultures were run on 15 % SDS-PAGE to find out the expression and solubility of the protein. The gel shows the full length protein where no any expressed protein is visible. Lane 1, 2, 3 are the pellets of control only rosetta cells, uninduced cells and induced cells respectively, lane 4, 5 and 6 represents the supernatant of control only rosetta cells, uninduced cells and induced cells respectively.



**Figure 2.4.2.3:** Agarose gel electrophoresis of the PCR amplified product of capsid protease domain of Aura virus (left panel). The cloned plasmid was screened for the presence of right construct and the restriction enzyme digestion shows the digested band at 500 bp which confirms the positive cloned plasmid (right panel).



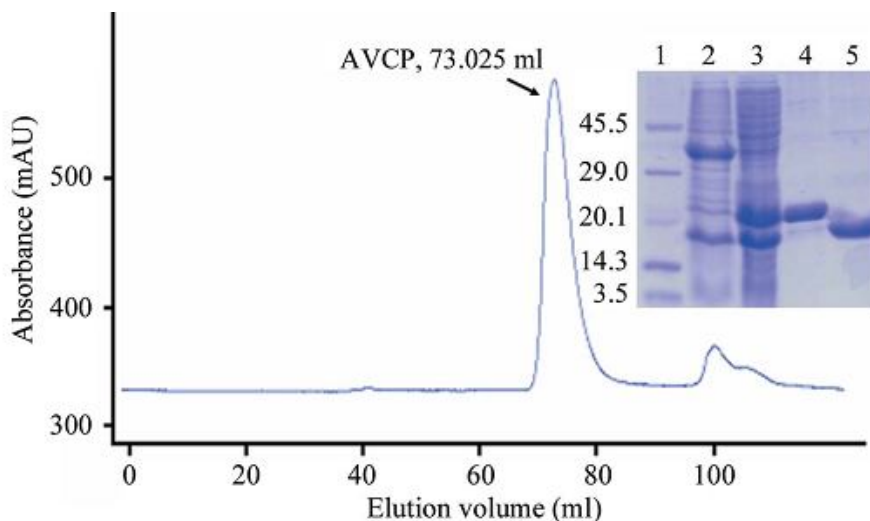
**Figure 2.4.2.4:** SDS-PAGE gel for the over-expression of AVCP. The supernatant along with the pellets were examined for the presence of soluble protein. The protein expression for the AVCP and the expressed protein band is visible with high solubility. Lane 1 shows the protein molecular weight marker; lane 2 contains the plane Rosetta cells having no plasmid. Lane 3 and 4 shows the uninduced pellet and supernatant respectively. Lane 5 and 6 presents the bands for induced cell pellet and supernatant respectively.

### 2.4.3 Purification of AVCP

The recombinant protein was purified using the IMAC (immobilized metal-affinity chromatography) method. The N-terminal His tag was cleaved by TEV protease and the protein was re-purified using a reverse Ni-NTA column. The flow through of the reverse Ni-NTA column containing AVCP without His tag was collected, concentrated and loaded onto the gel-filtration column. Size-exclusion chromatography, which was the last step of purification, provided a homogenous preparation of purified AVCP (Figure 2.4.3.1). Using a standard curve based on molecular-weight markers, the molecular weight of the major elution peak containing AVCP



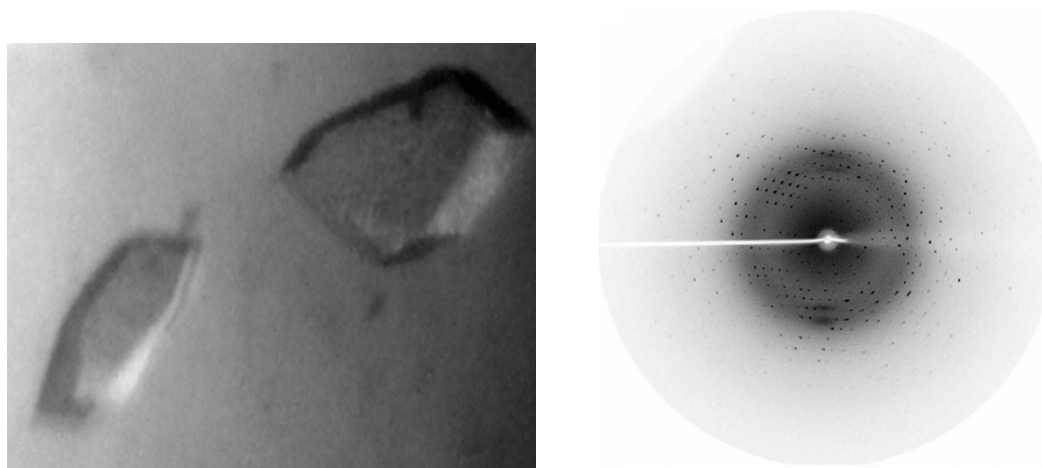
protein was calculated and was estimated to be approximately 17 kDa. This suggests that AVCP exists in a monomeric form. The estimated yield of pure protein was ~ 13 mg per liter of culture. The purified protein was concentrated to ~ 10 mg/ml and used for crystallization.



**Figure 2.4.3.1:** Gel-filtration profile and SDS-PAGE analysis of AVCP. Lane 1, Protein molecular-weight marker (kDa); lane 2, pellet containing insoluble protein fraction; lane 3, supernatant containing soluble protein fraction; lane 4, purified His-tagged AVCP; lane 5, purified AVCP without His tag.

#### 2.4.4 Crystal development and data collection

Crystals of AVCP were obtained in 15–20 days at 20 °C using 100 mM Bis-Tris buffer pH 6.5 and 28 % (w/v) Polyethylene glycol monomethyl ether 2000 as the precipitant (Figure 2.4.4.1). The crystals of AVCP were diffracted to 1.81 Å resolution with one molecule per asymmetric unit. The data-collection statistics for the native crystals are summarized in Table 2.1.



**Figure 2.4.4.1:** Crystals of Aura virus capsid protease (left). The longest dimensions of a typical crystal were ~ 50–100  $\mu\text{m}$ . Diffraction of AVCP crystals using in-house radiation (right panel). The resolution at the edge of the plate is 1.81 Å.

**Table 2.1:** Data collection and refinement statistics for AVCP in apo form

	AVCP-apo
Crystallographic Data	
Space group	<i>C2</i>
Cell dimensions (Å)	<i>a</i> =79.6 , <i>b</i> =35.2, <i>c</i> =49.5
Resolution range (Å)	50.0 – 1.81 (1.85–1.81)
(Last Shell)	
Completeness (%) (Last Shell)	90.2 (70.7)
<i>R</i> merge <sup>†</sup> (%) (Last Shell)	5.7 (25.5)
Mean <i>I</i> / $\sigma$ ( <i>I</i> ) (Last Shell)	22.6 (2.9)
No. of observed reflections	45053
No. of unique reflections	11410 (436)
(Last Shell)	
Molecules per asymmetric unit	1
Matthews coefficient (Å <sup>3</sup> Da <sup>-1</sup> )	2.0
Solvent content (%)	38.5
Multiplicity (Last Shell)	3.9 (2.1)
Refinement	
No. of Residues	152
Water molecule	127
<i>R</i> cryst (%)	17.6
<i>R</i> free (%)	23.15
Average <i>B</i> -factor (Å <sup>2</sup> )	23.54
r.m.s.d on bond lengths (Å)	0.012
r.m.s.d on bond angles (Å)	1.567
Ramachandran plot (%)	
Preferred	98.0
Allowed	2.0
Outliers	0.0

<sup>†</sup>  $R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - [I(hkl)]|}{\sum_{hkl} \sum_i I_i(hkl)}$ , where  $I_i(hkl)$  is the *i*th observation of reflection *hkl* and  $[I(hkl)]$  is the weighted average intensity for all observations *i* of reflection *hkl*.

The crystals belonged to the monoclinic space group  $C2$ , with 90.2 % completeness and an Rmerge of 5.7 %. The crystal structure of the capsid protease from Aura virus not only assists in structure-based drug design but also highlights the differences between alphavirus members that govern viral RNA-packaging and virus-assembly processes.

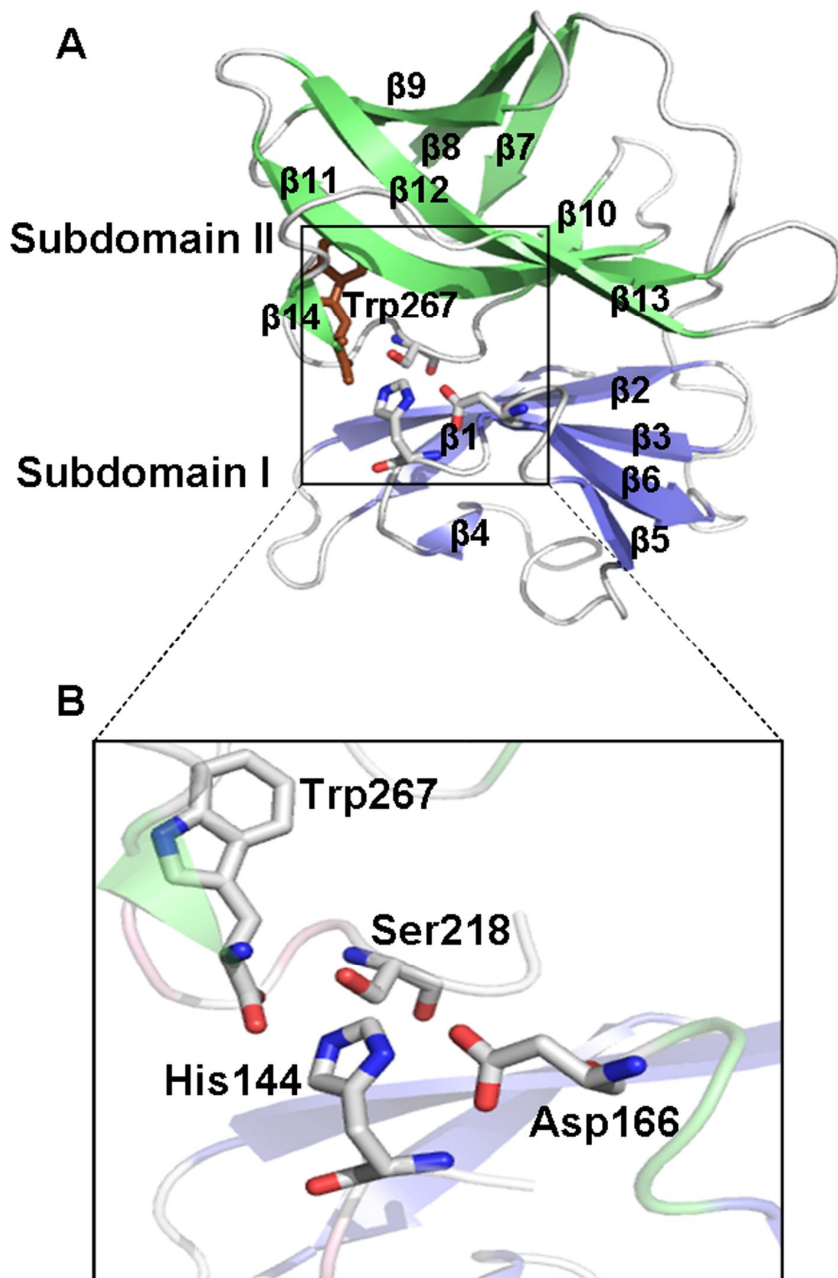
The crystal structure of AVCP was determined using molecular replacement method as described in Materials and Methods. The structure of AVCP was refined to 1.81 Å resolution with Rcryst value of 17.6 % and Rfree value of 23.1 %. The data collection and refinement statistics are presented in Table 2.1. The crystal belongs to the space group  $C2$  with cell dimensions  $a = 79.6$  Å,  $b = 35.2$  Å,  $c = 49.5$  Å and  $\alpha = 90.00$   $\beta = 102.34$   $\gamma = 90.00$ . The crystal is having a Matthews coefficient of  $2.0$  Å<sup>3</sup> Da<sup>-1</sup> and the solvent content of 38.5 %. According to the Ramachandran plot analysis, 98 % of the residues are found in the most preferred region. 152 out of 158 residues have well defined electron density while electron density is missing at the N-terminus for the first six residues (110–115) in AVCP. However, these residues were present in the cloned sequence. These residues were found to be very important in the interaction between the two monomers of crystallographic dimer in SCP (Choi *et al.*, 1996).

### 2.4.5 Overall structure of AVCP

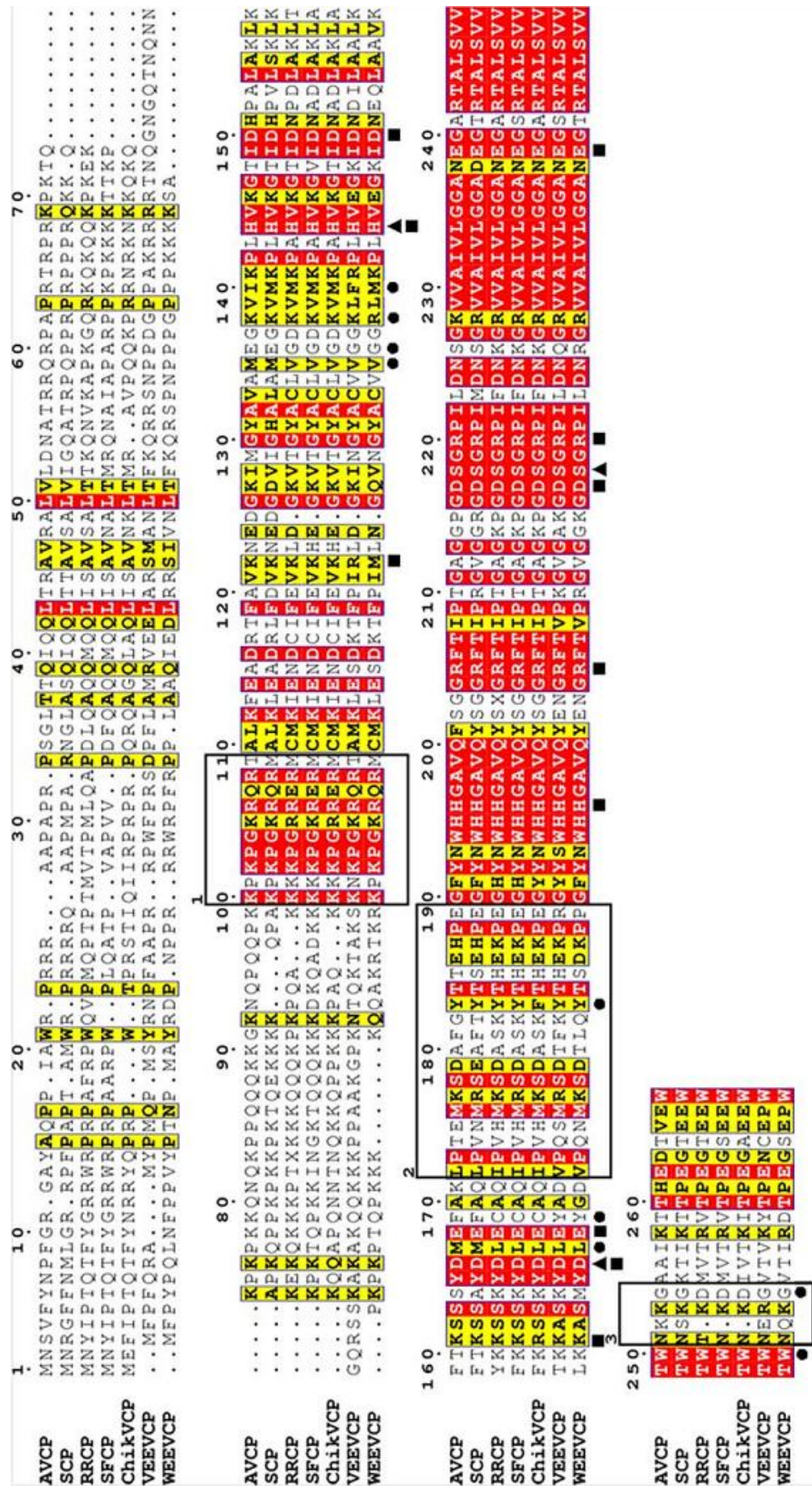
The overall fold of AVCP (C-terminal domain) consists of two  $\beta$ -barrel sub-domains similar to crystal structures of other alphavirus CPs (Choi *et al.*, 1997; Choi *et al.*, 1991). The sub-domains I and II consist of six and eight  $\beta$ -strands respectively, joined via a long linker loop region comprising of eighteen residues (Leu172-Glu189) (Figure 2.4.5.1). The twisted anti-parallel  $\beta$ -strands of individual sub-domains [ $\beta$ 1 (118–122),  $\beta$ 2 (128–135),  $\beta$ 3 (138–142),  $\beta$ 4 (148–149),  $\beta$ 5 (160–162),  $\beta$ 6 (167–171) of sub-domain I and  $\beta$ 7 (190–194),  $\beta$ 8 (197–202),  $\beta$ 9 (205–209),  $\beta$ 10 (221–223),  $\beta$ 11 (229–238),  $\beta$ 12 (242–250),  $\beta$ 13 (256–259) and  $\beta$ 14 (265–266) of sub-domain II] form the Greek key motif that is a characteristic feature of chymotrypsin-like serine proteases. Six salt bridges confined to individual sub-domains are reported which might be responsible for compact structure and stability. Three salt-bridges (Asp150-Lys122, Glu168-Lys161 and Asp166-His144) are confined to sub-domain I whereas other three Asp217-Arg220, Asp217-His196 and Glu239-Arg205 are confined to sub-domain II (Figure 2.4.5.2). The salt bridges help in stabilizing the structure of the protein.

The overall structure resembles the typical chymotrypsin like structure having the Greek key motif. However, some differences include the deletion of calcium binding loop and

methionine loop in AVCP which are the important part of the structure of chymotrypsin and play role in the catalysis. Also, there is no disulphide bond present in the structure in contrast to chymotrypsin or  $\alpha$ -lytic protease.



**Figure 2.4.5.1: Structure of Aura Virus Capsid Protease.** (A) Overall structure of the AVCP with two  $\beta$ -barrels of subdomains I and II colored in blue and green respectively. The catalytic triad residues and Trp267 are shown in sticks; (B) Close-up view of the active site shows catalytic triad composed of Ser218, Asp166 and His144 along with the carboxy-terminal Trp267 approaching the active site.



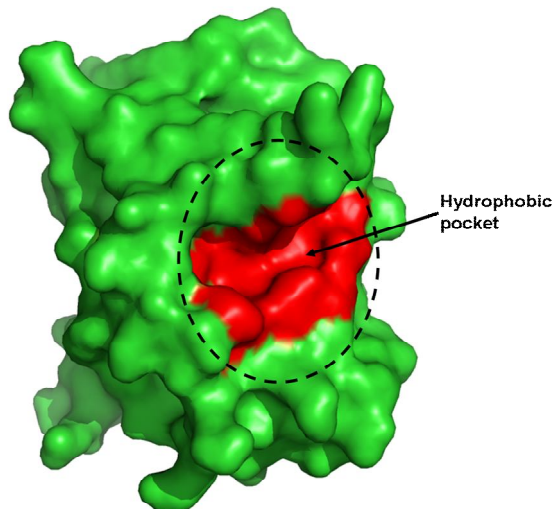
**Figure 2.4.5.2: Multiple sequence alignment of AVCP with CPs from other alphaviruses.** The conserved residues are shown in red background. The circles under the amino acids indicate the hydrophobic pocket residues interacting with the dioxane molecule while the triangles denote the catalytic triad residues. The residues forming salt bridges are denoted by squares below the residues. The motif responsible for specificity in encapsidation is highlighted in rectangular box 1. Interdomain flexible loop residues are shown in box 2 whereas the flexible loop separating the two pockets for E1 and E2 binding is shown in box 3. The CP sequences used for alignment are: AVCP, SCP, Ross River virus CP (RRCP), SFCP, Chikungunya CP (CHIKVCP), Venezuelan equine encephalitis virus CP (VEEVCP) and Western equine encephalitis virus CP (WEEVCP).

Alphavirus CP possesses cis-autoproteolytic activity for its cleavage from the rest of the structural polyprotein during the viral infection process (Hahn *et al.*, 1985). Once cleaved, the mature CP does not cleave other CPs further in trans due to the binding of the C-terminal Trp267 residue at the P<sub>1</sub> position near the active site (Tong *et al.*, 1993; Choi *et al.*, 1991). Due to conserved structural characteristics, the active site architecture in all members of the serine protease family appears to be quite similar. As a conserved feature of all the members of serine protease family, AVCP contains a catalytic triad formed of His144, Ser218 and Asp166 residues. The triad is located in the cleft between the  $\beta$ -barrel sub-domains (Figure 2.4.5.1). The backbone amides of Ser218 (the active site residue) and Gly216 form the oxyanion hole and act as donor of backbone hydrogen bond. Ser218 acts as nucleophile in the protease reaction. Trp267 remains bound to the active site through interacting with Ser218.

AVCP appears to be monomeric both in the crystal structure as well as in solution as observed by size-exclusion chromatography. In the crystal structures of SCP and SFCP, a crystallographic dimer with similar dimeric interface has been reported (Choi *et al.*, 1997; Choi *et al.*, 1991). Interestingly, fit of SFCP crystal structure in the cryo-EM density map suggests that molecular contacts similar to the dimer interface interactions are present in the virions. However, AVCP does not show the dimerization property even though the residues involved in crystallographic dimer formation in SCP are conserved with the sequence of AVCP (188-Pro-Glu-Gly-Phe-Tyr-Asn-193, Gly197, Ala198) (Figure 2.4.5.2). In addition to this, a single mutation at conserved F188G in SCP resulted in the disruption of the crystallographic dimer but had no effect on virus replication (Choi *et al.*, 1996; Tong *et al.*, 1993). Thus, the biological functions of the conserved dimer interface residues in the virus core structure formation need to be confirmed by additional mutational and structural studies.

On the other side of the active site, there is one more pocket present, which is named as the hydrophobic pocket (Figure 2.4.5.3) and is involved in virus budding as the E2 glycoprotein binds to the capsid through this hydrophobic pocket only. The hydrophobic pocket is surrounded by the

hydrophobic residues to make it compatible for binding to the E2 protein. The amino terminus residues of the protease domain of CP also have been proposed to bind to this pocket which mimics the binding of E2 glycoprotein cytoplasmic tail residues (Choi *et al.*, 1996).



**Figure 2.4.5.3:** The surface view of the crystal structure of AVCP showing the hydrophobic pocket in red.

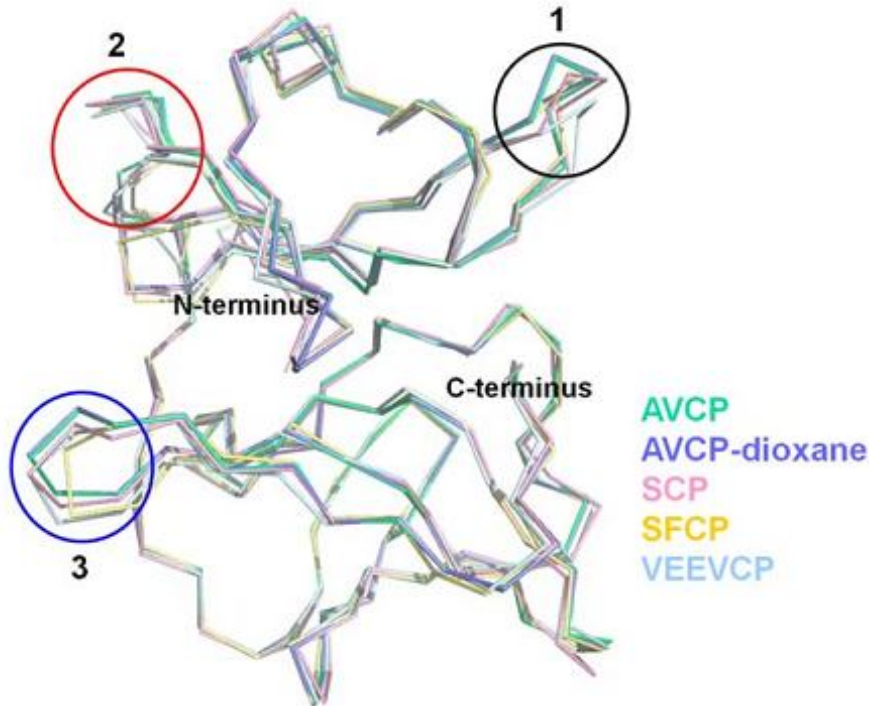
### 2.4.6 Comparative analysis of AVCP with capsid proteases from different alphaviruses

In order to attain a detailed comparative analysis, the superposition of AVCP with CPs from other alphaviruses was performed. The homology search for AVCP using blastp tool of National Center for Biotechnology Information (NCBI) (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) against the PDB database showed the hits of crystal structure of alphavirus capsid proteins with range 58 % to 79 % sequence identity. However, when the 3D coordinates of AVCP were used to search structural homologs using DALI server (Holm & Rosenström, 2010), the search results showed the most structurally conserved proteins of serine protease family including flavivirus NS3 protease, elastase,  $\alpha$ -lytic protease and chymotrypsin. Structural similarity between AVCP and flavivirus NS3 protease indicates that these may have evolved from a common chymotrypsin-like serine protease. Similarly, divergent evolution of alphaviruses and flaviviruses from a common ancestor has been suggested based on the structural homology between fusion glycoproteins E1 and E from alphaviruses and flaviviruses respectively (Kielian, 2002; Lescar *et al.*, 2001).

The  $\beta$ -barrel domain is well conserved among serine protease family; however, the size of the  $\beta$ -sheets varies. Superimposition of the apo form of AVCP with CP structures from other alphaviruses [SCP (RMSD = 0.481 Å), SFCP (RMSD = 0.468 Å) and VEEVCP (RMSD = 0.540 Å)] illustrates a high degree of structural similarity. Most notable variations are seen within the three loop regions (Figure 2.4.6.1). In SFV, fitting of CP crystal structure into cryo-EM density

## Chapter 2: Crystal structure of Aura virus capsid protein and insight into the capsid-glycoprotein interaction

suggests that these loop regions (SFV residues 125-128, 171-181, 252-255) are involved in CP-CP contacts in virions (Choi *et al.*, 1997). Particularly, AVCP does not possess any helix in the structure, whereas CPs of other alphaviruses possess a  $3_{10}$  and/or an  $\alpha$ -helix. The inter-domain loop region connecting the  $\beta$ -sheets is highly flexible and adopts different conformations (Circle 2, Figure 2.4.6.1). Sequence comparison shows that this inter-domain loop shares relatively less conserved region with respect to overall sequence identity (Box 2, Figure 2.4.5.2). The C-terminus of AVCP terminates with a short  $\beta$ -strand consisting of Val265-Glu266 residues that has been proposed to keep the C-terminal Trp residue in a static position near the active site (Tong *et al.*, 1993). This C-terminus Trp occupies the same spatial position in all alphaviruses and resides near the active site needed for autoproteolysis.



**Figure 2.4.6.1: Structural alignment of CPs from different alphaviruses.** Structure alignment of both apo (green) and complex (blue) form of AVCP with SCP (pink), SFCP(yellow) and VEEVCP (cyan) with PDB IDs: 1KXA, 1VCP and 1EP5 respectively showing structural variation. Three circles show loop regions of high variability that include inter-domain flexible loop region (Circle 2) and flexible loop that separate the two pockets responsible for binding to E1 and E2 (Circle 3).

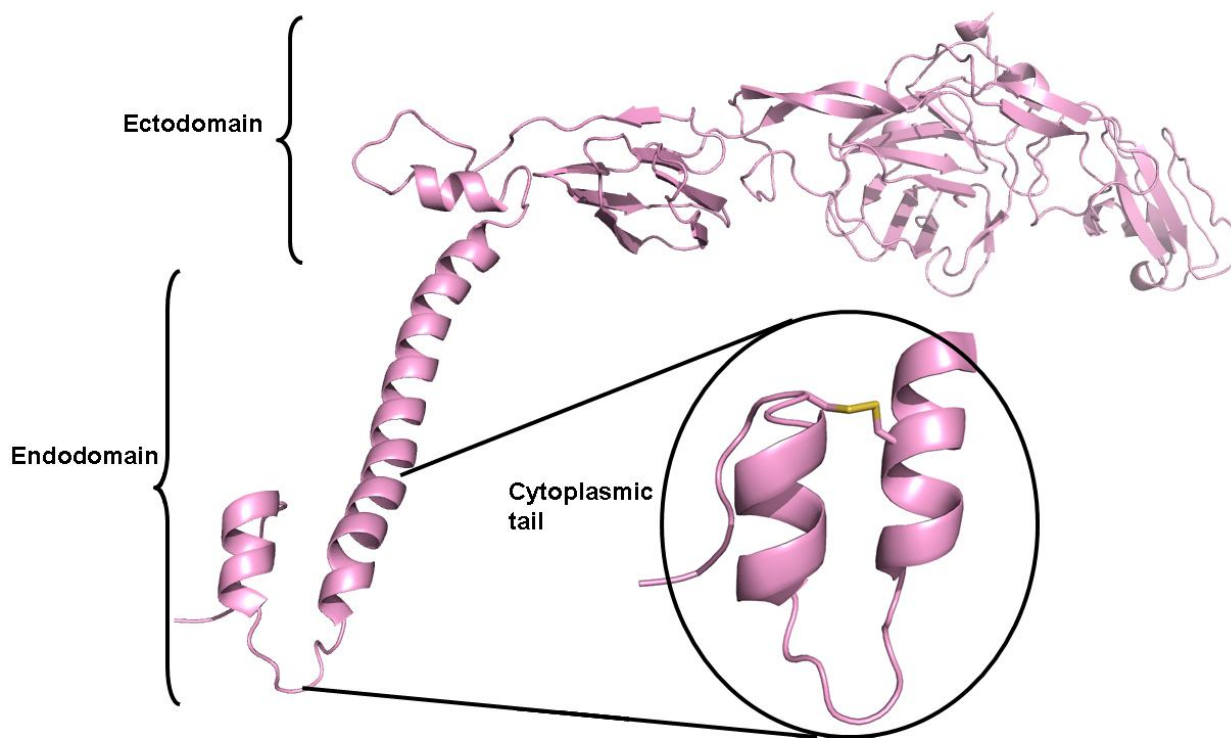
Apart from structural comparison, sequence alignment also provides interesting information regarding encapsidation of alphaviruses. The N-terminal motif Lys97-Met106 of SCP (corresponding to Lys100-Thr109 of AVCP) was found to be responsible for specific encapsidation as this motif has been shown to be required for specificity of genomic RNA



encapsidation (Owen & Kuhn, 1996). However, even though the motif is highly conserved in all alphaviruses, Aura virus encapsidates both genomic RNA and subgenomic RNA in virus particles (Rümenapf *et al.*, 1994) (Figure 2.4.5.2). Therefore, the possible explanations for encapsidation of the subgenomic RNAs by Aura virus could be that the subgenomic RNAs of Aura virus may contain a packaging signal. The identification of a packaging signal, if any, in Aura virus subgenomic RNA needs to be investigated further.

### **2.4.7 Homology Modeling of E1 and E2 glycoproteins**

In the process of investigating the CP-glycoprotein interaction, the homology models for both E1 and E2 glycoproteins were prepared by taking VEEV glycoproteins structure as template. The homology model of Aura virus E2 glycoprotein is shown in Figure 2.4.7.1.

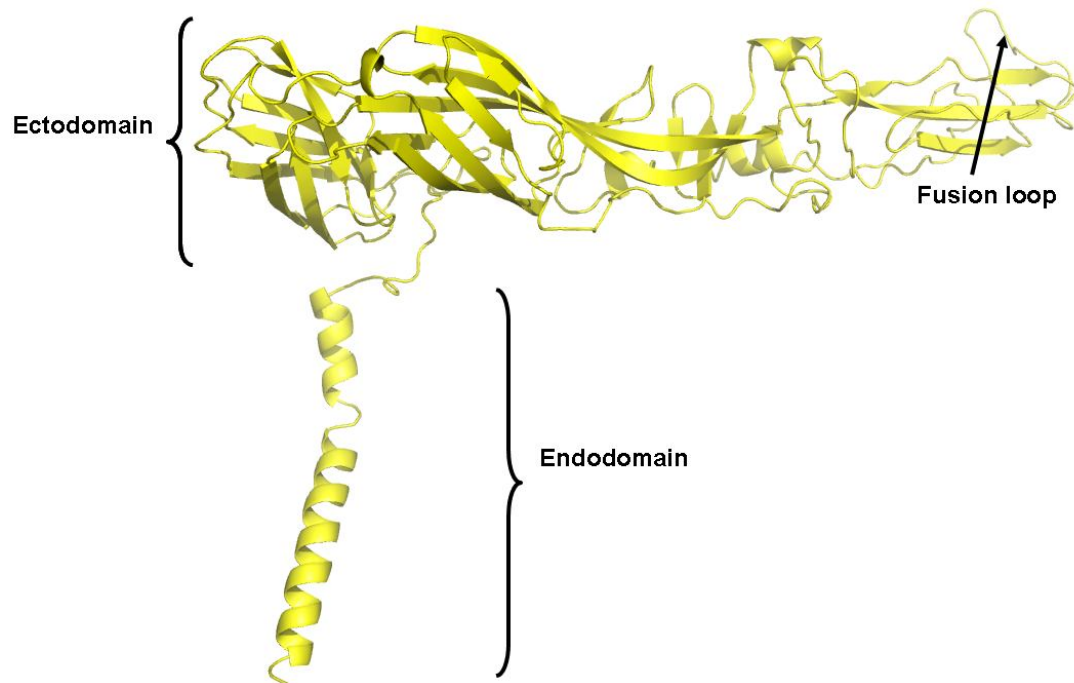


**Figure 2.4.7.1:** The cartoon representation of the E2 glycoprotein showing the ectodomain as well as endodomain. The endodomain includes the cytoplasmic tail for which the zoom-in view is shown. The zoom-in view of the cytoplasmic tail shows helix-loop-helix motif having a disulphide bridge (yellow color) joining the two helices.

The E2 glycoprotein consists of two domains: ecto and endo domains. The ectodomain contains several subdomains which include the sites for N-linked glycosylation as well as the epitopes for mAB binding.

## Chapter 2: Crystal structure of Aura virus capsid protein and insight into the capsid-glycoprotein interaction

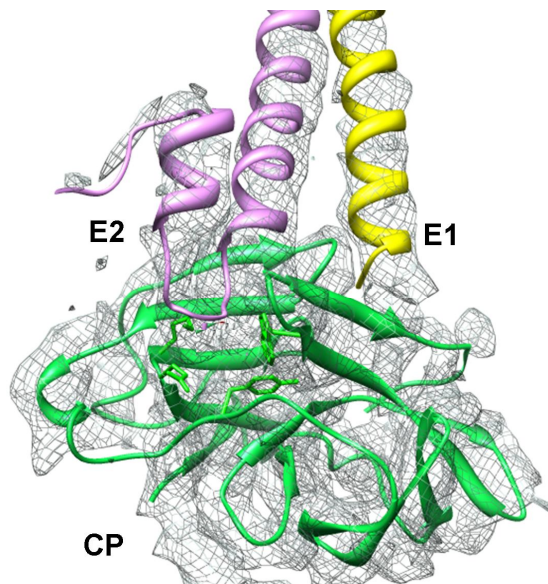
Likewise, the E1 glycoprotein also contains ecto and endo domains. The fusion loop (involved in the fusion during virus entry and gets exposed with the change in pH from neutral to acidic) is present in the ectodomain. Unlike the E2 glycoprotein, E1 endodomain has only the helical structure which consists of two consecutive helices. These helices are separated by a kink. The model of the E1 glycoprotein including both the domains is shown in Figure 2.4.7.2.



**Figure 2.4.7.2:** The homology model of the E1 glycoprotein is shown in the cartoon view. The two domains are present in the E1 protein: endo and ecto domain. The fusion loop is present in the ectodomain of the protein. The endodomain is present in the form of long helix.

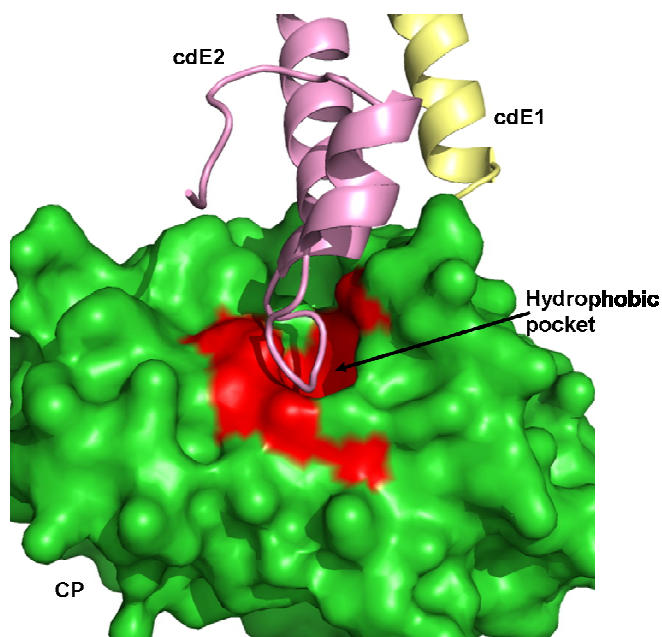
### 2.4.8 Interaction analysis of E1 and E2 glycoproteins with the capsid protease

The recent cryo-EM structure of VEEV opened up the possibility to explore the interaction patterns of glycoproteins with CP (Zhang *et al.*, 2011). The VEEV structure revealed that the cytoplasmic tail of E2 interacts with the hydrophobic pocket of capsid. Furthermore, Tang *et al.*, 2011 elucidated the interaction at molecular level by the cryo-EM structure of Sindbis virus and identified by mutagenesis that Arg393 and Glu395 residues of cdE2 and Tyr162 and Lys252 of capsid are the key residues for interaction (Tang *et al.*, 2011). However, the exact interaction is not yet clear. To throw light on the mode of interaction, we fitted the crystal structure of AVCP and homology models of the cytoplasmic tails of E1 and E2 trans-membrane glycoproteins into the cryo-EM density map of VEEV (Figure 2.4.8.1). The superposition of Aura virus E1 and E2 cytoplasmic tails to that of VEEV shows RMSD of 0.206 Å and 0.578 Å respectively.



**Figure 2.4.8.1:** Representation of the crystal structure of AVCP-dioxane complex and the homology models of E1 and E2 glycoproteins fitted into the cryo-EM electron density map of VEEV (EMDB ID: 5275). AVCP, E1 and E2 are shown in green, yellow and pink colors respectively; the cryo-EM density map is denoted by gray.

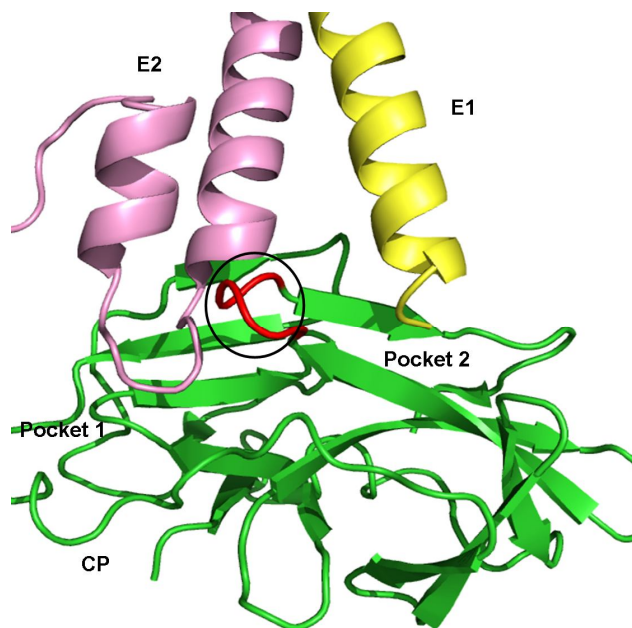
The C-terminus of E2 is highly conserved in all alphaviruses and possesses a tightly folded helix-loop-helix motif that makes a globular structure to fit into the hydrophobic pocket of AVCP (Figure 2.4.8.2).



**Figure 2.4.8.2:** The cytoplasmic domain of E2 glycoprotein forms the helix-loop-helix motif. The loop of this motif interacts with the hydrophobic pocket of the CP. The E1 protein also comes close to the capsid which might be involved in CP-glycoprotein interaction. The E2, E1 and CP are shown in pink, yellow and green colors respectively. The hydrophobic pocket is presented in red.

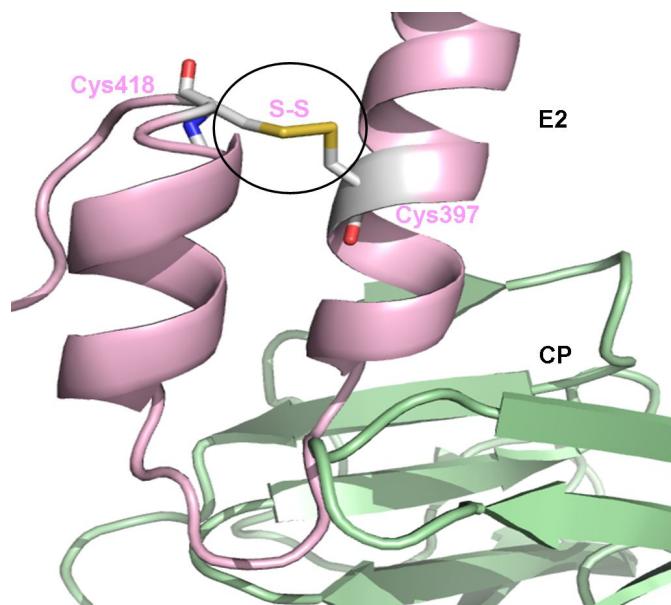
## Chapter 2: Crystal structure of Aura virus capsid protein and insight into the capsid-glycoprotein interaction

Furthermore, the cryo-EM structure also revealed the mode of E1-capsid interaction. E1 glycoprotein in Aura virus is expected to interact in a small depression in the CP (P2 pocket), which is separated from the E2-interacting hydrophobic pocket (P1 pocket) by a small flexible loop stretching from Asn251 to Ala255 of AVCP (Figure 2.4.6.1, Figure 2.4.8.3). Although the exact mode of interaction is not clear, it can be assumed that the E1 cytoplasmic tail extends towards the capsid with limited polar interactions with the E2 cytoplasmic tail.



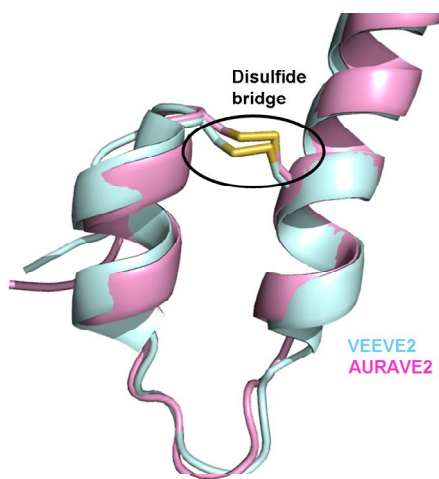
**Figure 2.4.8.3: Cartoon view of the AVCP-glycoproteins interaction.** Binding of Aura virus E2 and E1 cytoplasmic tails to different pockets (P1 and P2 respectively) found in close proximity at the surface of AVCP. E2 shows the helix-loop-helix structural architecture and E1 has a helical structure that interacts with AVCP. The loop separating both interacting pockets is shown in red color.

Additionally, Jose *et al.*, 2012 recently demonstrated that Cys residues in cdE2 are critical for the budding process of alphaviruses, as these have been proposed to undergo palmitoylation to develop a hydrophobic region that facilitates interaction with capsid (Ivanova & Schlesinger, 1993). Structural elucidation reveals that four Cys residues i.e. Cys397, Cys416, Cys417 and Cys418 are present at the C-terminal tail of Aura virus E2. Interestingly, Cys397 is found to make a disulfide bridge with Cys418 to establish a stable C-terminal lobe structure (Figure 2.4.8.4).

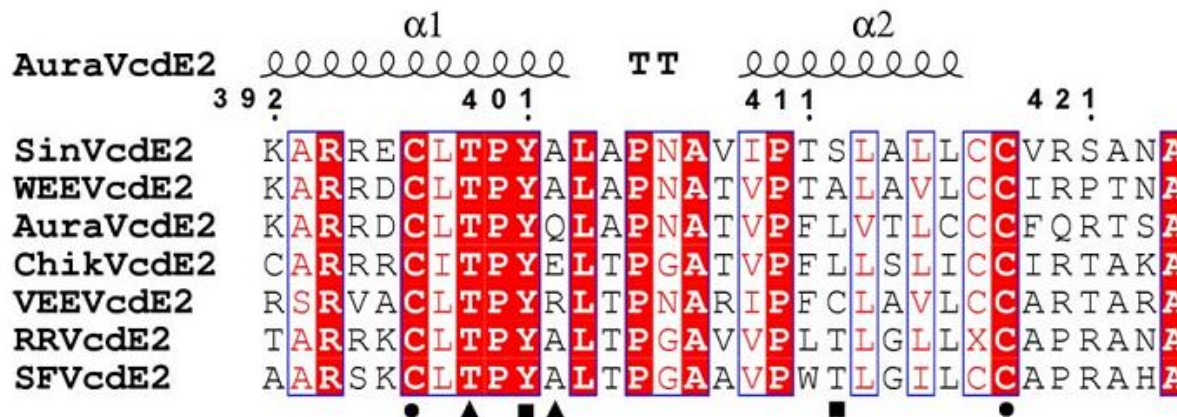


**Figure 2.4.8.4:** The cytoplasmic tail of E2 glycoprotein from Aura virus showing the Cys-Cys disulfide bond in between helix-loop-helix motif of E2. The disulfide bond is highlighted in circle in yellow color.

Moreover, we have also analyzed the cryo-EM structure of VEEV and found that the VEEV cdE2 also contains a disulfide bond at the same position (Figure 2.4.8.5). This disulfide bridge is expected to be formed in cdE2 of all alphaviruses as both the disulfide forming cysteine residues are conserved at the corresponding positions (Figure 2.4.8.6). Furthermore, the proposed structural role of cdE2 disulfide bridge is supported by mutational data of Sindbis virus cdE2 Cys417 mutation (Cys418 in Aura virus) which shows that Cys417Ala mutation slows down the virus replication and decreases the infectious virus yield (Ryan *et al.*, 1998).

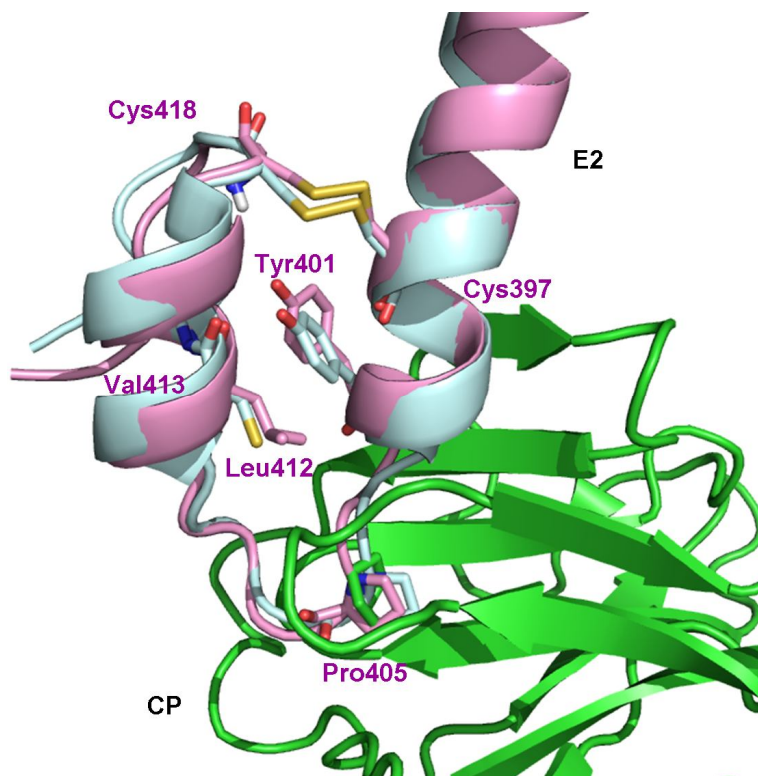


**Figure 2.4.8.5:** The structural alignment of the cytoplasmic domain of E2 glycoproteins from AURAV and VEEV. The disulfide bond is formed by the two cysteine residues from the opposite helices of the helix-loop-helix motif. The disulfide bond is presented as sticks in yellow color. The E2 from VEEV and AURAV are shown in blue and pink color respectively.



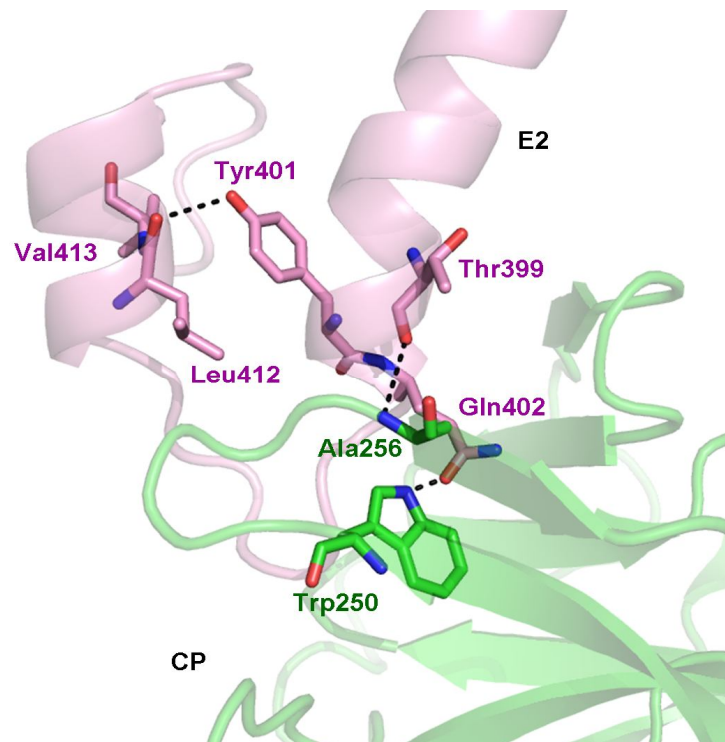
**Figure 2.4.8.6: Multiple sequence alignment of the cytoplasmic domain of E2 (cdE2) from different alphaviruses.** The secondary structure of 33 residues from Aura virus cdE2 displays a helix-loop-helix structure. Highly conserved residues are in red background. The circles under residues denote the disulfide bridge forming Cys residues, squares indicate the residues responsible for polar interactions within the E2 helix-loop-helix motif and the triangles represent the residues of E2 making polar interactions with CP. cdE2 sequences used for alignment are: Sindbis (SinVcdE2), Western equine encephalitis (WEEVcdE2), Aura (AuraVcdE2), Chikungunya (ChikVcdE2), Venezuelan equine encephalitis (VEEVcdE2), Ross River (RRVcdE2) and Semliki Forest (SFVcdE2).

Structural analysis of the interactions between cdE2 and capsid reveals that the loop region of the helix-loop-helix motif of cdE2 docks into the hydrophobic pocket of AVCP (Figure 2.4.8.1 and 2.4.8.2). The pyrrolidine ring of Pro405 of the loop region in cdE2 makes hydrophobic contacts with the hydrophobic pocket of CP. The cryo-EM structure of VEEV also shows that the conserved proline of cdE2 occupies the same position as Pro405 of Aura virus cdE2 in the hydrophobic pocket of CP (Figure 2.4.8.7). These structural investigations suggest that Pro405 might play a crucial role in alphavirus budding process. Thus, the role of Pro405 in the budding process has been established through structure based analysis. Our observation is supported by mutational studies in Sindbis virus cdE2 where Pro404Ala substitution (corresponding to Pro405 of Aura virus cdE2) results in slow-growth phenotype. Also, this mutant was found to be more thermolabile than the parental wild type virus (Ryan *et al.*, 1998; Ivanova & Schlesinger, 1993). Also, this proline residue is conserved in all the members of alphavirus genus (Figure 2.4.8.6). Hence, it can be suggested that the proline residue from the loop region of cdE2 is directly involved in CP-glycoprotein interaction. The hydrophobic pocket creates the hydrophobic environment for proper binding of the proline to the pocket. Thus, it can be assumed that the budding process takes place by interaction of the CP-glycoprotein through binding of the proline from the loop region to the hydrophobic pocket of alphavirus CP.



**Figure 2.4.8.7:** Structural alignment of cdE2 predicted structure from Aura virus (pink) and VEEV (blue) showing Tyr401 residue oriented towards other helix away from CP and the disulfide bond (yellow) within the helix-loop-helix motif of E2. The Pro405 interaction with the hydrophobic pocket present on the surface of the AVCP crystal structure is also shown.

The conserved residue Tyr400 in Sindbis virus E2 was proposed to be responsible for E2-capsid interaction (Skoging *et al.*, 1996; Zhao *et al.*, 1994). Crystal structure of SCP showed the binding of N-terminal arm hydrophobic residues Leu108 and Leu110 in the hydrophobic pocket of the neighboring CP molecule (Choi *et al.*, 1991). This interaction is similar to the binding of the E2 residue Tyr400 with the hydrophobic pocket. Additionally, mutational analysis of Sindbis virus showed that for successful virus propagation only hydrophobic residues could substitute the conserved Tyr400 (Tyr401 in Aura virus) (Zhao *et al.*, 1994). Based on these observations, it was assumed that Tyr400, a conserved residue of cdE2 binds into the hydrophobic pocket of CP. However, in the generated model of Aura E2, Tyr401 makes a hydrogen bond with Leu412 and forms hydrophobic contacts with Leu412 and Val413 of the opposite helix. Hence, Tyr401 residue of cdE2 is oriented away from the hydrophobic pocket of capsid towards the other helix in helix-loop-helix structure (Figure 2.4.8.8). Thus, it depicts that Tyr401 may play a critical role in maintaining the helix-loop-helix structure of cdE2. In addition, few polar contacts are found to be involved in stabilizing the E2-capsid interaction. Thr399 and Gln402 of E2 glycoprotein show polar contacts with Ala256 and Trp250 of AVCP respectively (Figure 2.4.8.8).



**Figure 2.4.8.8:** The critical polar interactions between E2 and capsid as well as within the helix-loop-helix motif of cdE2 are shown in dotted lines. E2 residues are shown in pink while the AVCP residues are displayed in green. The interacting residues are shown as sticks.

The structure analysis reveals the presence of E2 residue Pro405 into the hydrophobic pocket of AVCP and the interaction of Tyr401 with the residues from the opposite helix. Even Tyr401 does not reach up to the hydrophobic pocket. Therefore, we can conclude that Tyr401 might not directly involve in CP-glycoprotein interaction through binding to the hydrophobic pocket. On the other hand, Pro405 bind to the hydrophobic pocket as determined through the cryo-EM structure of VEEV as well as the homology model of AURAV E2 fitted into the density map of VEEV. However, as described above, earlier studies suggest the role of Tyr400 (Sindbis residue) in the virus budding and can only be replaced by the hydrophobic residue only. It might be due to the disruption of the interaction of Tyr401 (Aura residue) with the residues on the opposite helix. Thus, the stabilization of the helix-loop-helix motif is highly essential for the interaction with the CP and finally for budding.

The hydrophobic environment, electrostatic interactions and the disulphide bonds, all are involved in the stabilization of the helix-loop-helix motif of the E2 glycoprotein cytoplasmic tail. The disruption in any of these interactions might lead to the defect in virus budding. Also, Pro405 has been established as the binding residue of the E2 protein with the hydrophobic pocket of AVCP. The mutation in any of these residues (Cys 397, Cys418, Tyr401 and Pro405) results in



either defective or slow growing virion (Jose *et al.*, 2012; Ryan *et al.*, 1998; Zhao *et al.*, 1994; Ivanova & Schlesinger, 1993). The structure analysis of these important residues in the process of virus budding has been demonstrated. The results indicate the direct involvement of Pro405 in the virus budding as it binds to the hydrophobic pocket of the CP. Other residues cysteines and tyrosine might help in stabilizing the helix-loop-helix motif of the E2 glycoprotein tail through disulphide bond formation, electrostatic and hydrophobic interactions with the residues from the opposite helix in the same motif. All these residues, Pro405, Cys397, Cys418 and Tyr401 are conserved among all the members of the alphavirus family (Figure 2.4.8.6). The E1 glycoprotein has also been found to play some role in budding through interaction with CP as its cytoplasmic tail approaches towards the pocket (other than hydrophobic pocket) of CP. The two pockets are separated by a flexible loop. However, more studies are required to get the complete knowledge of the interaction of CP and E1 glycoprotein.

### 2.5 Conclusion

The CP of alphaviruses performs several functions for virus life cycle including the proteolytic activity for structural polyprotein processing, interaction with the genomic RNA to encapsidate it, interaction with the other capsid to form the nucleocapsid assembly and interaction with the glycoproteins required for virus budding. Due to its multifunctional approach for the virus life cycle, it can be a good target for antiviral drug development against alphavirus infection. In this chapter, the highly important function of the CP to interact with the glycoproteins for virus budding has been studied. The mode of interaction of the glycoproteins with the capsid was investigated at the molecular level by comparing sequences and structures from different alphaviruses. The role of various conserved residues involved in CP-glycoprotein interaction has been proposed.

The interaction of glycoproteins with the capsid is crucial for alphavirus budding process. Recent cryo-EM studies revealed the overall interaction pattern of these proteins. In the present study, we have determined the crystal structure of AVCP and generated homology models of E1 and E2 from Aura virus. To investigate CP-glycoprotein interactions in more details, the AVCP crystal structure and homology models of E1 and E2 were fitted into the cryo-EM density map of VEEV. The structural elucidation and extensive interaction analysis revealed several interesting features. The cytoplasmic tail of E2 is predicted to contain a helix-loop-helix topology that may play an important role in CP-E2 interaction. Previous studies have suggested that residues Tyr401,

## **Chapter 2: Crystal structure of Aura virus capsid protein and insight into the capsid-glycoprotein interaction**

Cys397 and Cys418 in E2 are important for interaction of E2 with CP. Interestingly, comparative modeling of Aura virus E2 suggests that these residues play a major role in the stabilization of the helix-loop-helix topology of cdE2. Additionally, comparative modeling and fitting studies suggest that Pro405, a conserved residue present in the loop region of the helix-loop-helix motif of Aura virus interacts directly with the hydrophobic pocket of CP. The proposed role of various Aura virus cdE2 residues needs to be confirmed by mutational and crystal structural studies of capsid-cdE2 peptide complexes.

The E1 glycoprotein also interacts with the CP through its cytoplasmic tail. The interaction occurs with the pocket which is different from the hydrophobic pocket to which E2 binds. The two pockets are separated by a highly flexible loop. Thus, E1 might also help in virus budding through interaction with the CP. The structure of AVCP that elucidates the hydrophobic pocket, comparative modeling of glycoproteins and cryo-EM fitting studies that depict the binding pattern of CP-glycoproteins will define a valuable roadmap for developing a new strategy for antiviral drug discovery.

### 3.1 Abstract

This chapter demonstrates the crystal structure of the AVCP in complex with dioxane, piperazine and picolinic acid. The cryo-EM fitted structure (described in Chapter 2) has been analyzed using the dioxane bound complex structure. The position of the dioxane in the AVCP is assessed on the basis of the E2 glycoprotein tail bound to the hydrophobic pocket of AVCP. The three crystal structures bound to different ligands are analyzed and the comparative analysis has been performed. The docking studies with other potential antiviral compounds have also been performed for the AVCP structure.

The AVCP crystals were developed as described in Chapter 2. The complex formation was performed by the soaking of these native AVCP crystals into the buffer containing ligand. After soaking, the crystal was mounted and the data were collected to 1.98, 1.73 and 2.2 Å resolution for dioxane, picolinic acid and piperazine bound complex respectively. The space group and the cell parameters remain same with the native form for all the three complexes. The Matthews coefficient as well as the solvent content also remain same and found to be  $2.0 \text{ \AA}^3 \text{ Da}^{-1}$  and 38.5 % respectively. For all the complex forms, a single monomer per asymmetric unit is present.

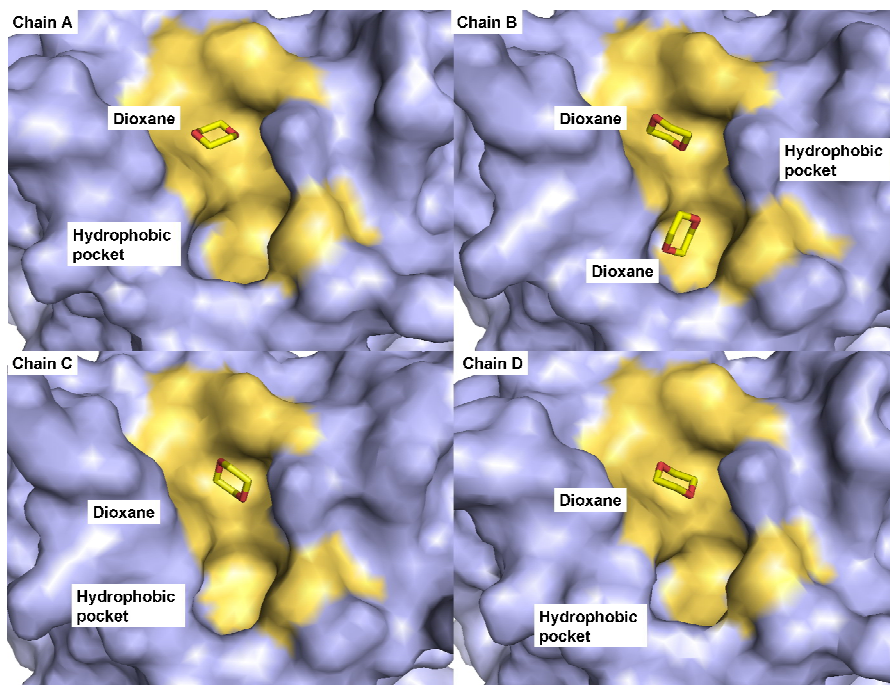
The overall structure of the native AVCP and AVCP complexes remains same and shows the similar “Greek key” motif. The dioxane molecule was found to be present in the hydrophobic pocket of AVCP. This pocket is same to which the E2 glycoprotein binds that leads to budding of the virus. The basis for disruption of these interactions by binding of dioxane in the hydrophobic pocket of CP has been proposed. In the AVCP-dioxane complex structure, dioxane occupies the hydrophobic pocket of CP and structurally mimics the hydrophobic pyrrolidine ring of Pro405 in the loop region of E2. As Pro405 of E2 glycoprotein is directly involved in binding to the hydrophobic pocket of AVCP, the dioxane molecule (which is present exactly at the same position) would competitively inhibit the binding of proline to the pocket. Thus, dioxane disrupts the E2 glycoprotein and CP interaction which would halt the virus budding process. Other compound piperazine is also present at the same position where dioxane was observed, however, the picolinic acid does not bind to the hydrophobic pocket of AVCP. The picolinic acid is found to be present at the surface of the protein. The compounds used for the docking were having two ring structures which get fit into the two grooves of the hydrophobic pocket. The docked structures were compared for their binding affinities with the hydrophobic pocket.

### 3.2 Introduction

The CP is a multifunctional protein and performs autoproteolysis, RNA binding, nucleocapsid assembly and budding process for virus infection. For the budding process, the CP interacts with the cytoplasmic domain of the glycoproteins (Lee *et al.*, 1996). This interaction has been studied in detail to find out the exact mechanism of budding. The E2 glycoprotein interacts with the hydrophobic pocket of the CP. Similarly, the amino-terminus of the CP was found to interact with the neighboring capsid molecules in the same manner as E2 glycoprotein binds. The residues Leu108 and Leu110 in SINV create the hydrophobic environment and binds to the hydrophobic pocket imitating the binding of Tyr400 and Leu402 of E2 tail residues. However, the structural studies on SFCP show no contact of the N-terminal arm with the hydrophobic pocket (Choi *et al.*, 1997). Thus, in SFCP, the hydrophobic pocket remains empty. Due to this difference in the binding mode, some structural differences appear between SCP and SFCP. Tyr180 and Trp247 residues of the CP from SINV and SFV were showing different structural conformations. However, this conformational change can also be the result of the difference in the structure of the two different viruses from the same family.

In order to get the clear picture for the different conformations of the two crystal structures, the attempts have made to crystallize the SCP without the N-terminal arm (residues 114-264). The protein was crystallized successfully and the conformation remains same with the previously determined structure with the N-terminal arm. However, the crystal structure of CP from Sindbis virus was found to contain the solvent-derived dioxane in its hydrophobic pocket (Lee *et al.*, 1998). This suggested that dioxane or similar molecules may be able to enter the pocket. The hydrophobic pocket contains two dioxane molecules in its two separate depressions in one chain of the CP. However, the rest of the three chains are having a single dioxane molecule corresponds to one depression in the pocket. Also, the conformation of the hydrophobic pocket differs between different chains of CP.

Furthermore, the two dioxane molecules that occupy different grooves in the hydrophobic pocket were considered to be bound in the same way as E2 residues (Tyr-X-Leu) or N-terminal arm residues (Leu-X-Leu) bind. The two hydrophobic residues might occupy the same position that the dioxane molecule occupies. However, other chains consist of just one dioxane molecule and the other groove remains empty (Figure 3.2.1). The exact mechanism of inhibition by the dioxane molecule is not clear yet.



**Figure 3.2.1: Hydrophobic pocket of different chains of SCP (PDB ID: 1WYK).** Chain B is having two dioxane molecules in different grooves of the hydrophobic pocket while all other chains contain only one dioxane. The pocket dimensions vary in different chains of the SCP showing its flexibility. The hydrophobic pocket forming residues are shown in yellow color surface.

It was assumed that the binding of the dioxane in the hydrophobic pocket might inhibit the binding of the E2 glycoprotein to the capsid and thus preventing virus budding. That opens up the possibilities of the design of dioxane based inhibitors for targeting the budding process of alphaviruses. However, the compound must have a high affinity towards the hydrophobic pocket. In order to perform the inhibition studies, initially four compounds were selected on the basis of the docking results with the crystal structure of SCP (Kim *et al.*, 2005). The compounds were tested against SINV having the firefly luciferase expressing gene as well as were tested for the cytotoxicity. Four compounds were designed initially to perform the inhibition studies. The inhibition experiment was performed with the firefly luciferase reporter gene for the analysis of virus infection in the BHK (Baby Hamster Kidney) cells. The comparative antiviral activities and the cytotoxicity of the compounds were analyzed (Kim *et al.*, 2005).

Afterwards, a number of dioxane based synthetic antiviral compounds were designed, synthesized and tested for the antiviral activity based on the crystal structure of SCP containing dioxane in the hydrophobic pocket (Kim *et al.*, 2007). However, no drug has yet been approved against alphavirus infection. Hence, it is important to develop antivirals against alphaviruses where budding process can be a potential target. Dioxane derivatives were considered as potential drugs

for future antiviral development. However, the exact mechanism of disrupting the CP-glycoprotein interaction is not clear. Therefore, more studies are required to get the knowledge of the dioxane disruption mechanism.

### 3.3 Methodology

#### 3.3.1 Strategy used for the study

The strategy used for the study of dioxane and similar molecules as antivirals is shown in Figure 3.3.1.1.

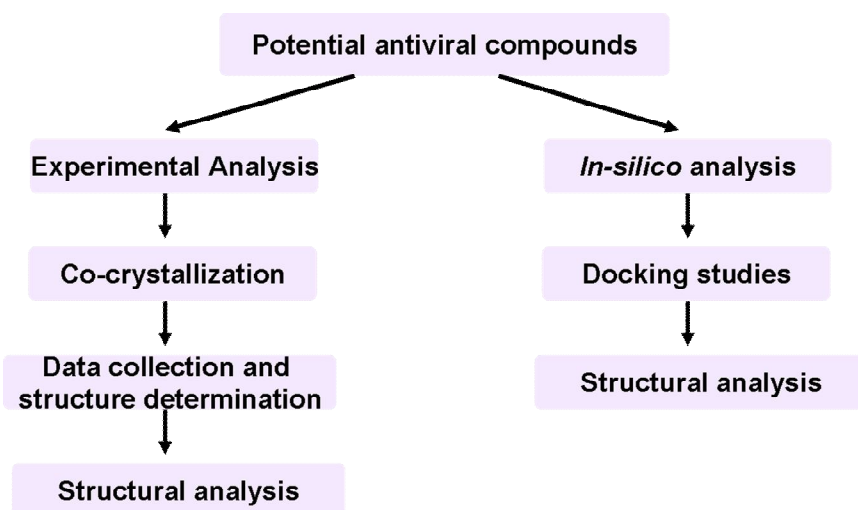


Figure 3.3.1.1: The schematic representation of the strategy used for the study.

#### 3.3.2 Co-Crystallization and Data Collection

The purification and crystallization of the AVCP was done as described in Chapter 2. Further, the native protein crystals were soaked in mother liquor containing 10 mM dioxane to obtain the AVCP-dioxane complex. After soaking the crystal, it was mounted and the data were collected. For the preparation of the complexes with piperazine and picolinic acid, the native AVCP crystals were produced in 0.1 M Bis-Tris pH 6.5, 22 % PEG (polyethylene glycol) monomethyl ether 2,000. The native crystals were soaked in mother liquor containing 10 mM piperazine for approximately 15 minutes. 17 % ethylene glycol was used as a cryoprotectant. The AVCP-picolinic acid complex was formed by soaking the native protein crystal in reservoir buffer containing picolinic acid for 5 minutes and mounted for data collection. For cryo-protection of the crystal, 17 % ethylene glycol was used. The data collection and the data processing were done as described in previous Chapter 2.

### 3.3.3 Structure solution and refinement

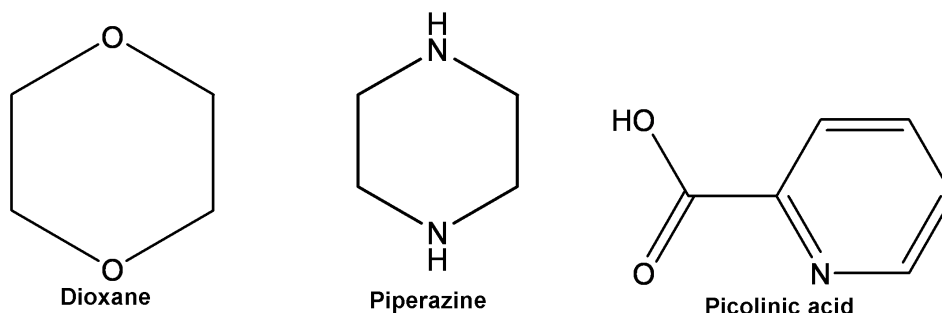
The structure determination was done using the molecular replacement method by using the crystal structure of native AVCP as the template. Geometric restraints for the ligands were generated using ProDRG, an automated topology generation server (Schuttelkopf & van Aalten, 2004). The generated ligand molecules were closely fitted into  $2F_o - F_c$  electron density and then included in following fitting and refinement cycles. The program COOT was used for model building and Refmac5 was used for refinement (Emsley & Cowtan, 2004; Murshudov *et al.*, 1997). A repetitive cycle of both refmac5 and coot was run to get the final refined model. The structure was evaluated by using PROCHECK server (Laskowski *et al.*, 1993). The figures were prepared using PyMol visualization tool (DeLano, 2002).

### 3.3.4 Molecular Docking and Analysis

The docking analysis was performed using GLIDE v 5.5 software (Grid-based ligand docking with energetics) by Schrödinger. The software was run on Red Hat enterprise Linux 5 operation system installed on HPxw8400 workstation (Cross *et al.*, 2009). Firstly, the protein was prepared in protein preparation wizard of Maestro for the addition of hydrogens, the assignment of bond orders and the energy minimization of the structure up to RMSD value of 0.3 Å. For the minimization, OPLS2005 force field was used. The compounds were generated using PubChem (<http://www.ncbi.nlm.nih.gov/pccompound>) and Chemdraw. The ligands were prepared using Maestro's ligand preparation module. The amino acid residues of the hydrophobic pocket were selected and Glide receptor grid was prepared. The extra precision mode was used for the docking of the ligands (Friesner *et al.*, 2004). The docked structures were analyzed by evaluating the glide energy and glide score. The visual inspection of the structures was done using PyMol visualization tool. After evaluation of the structure, the best pose was selected for all the ligands (DeLano, 2002).

## 3.4 Results and Discussion

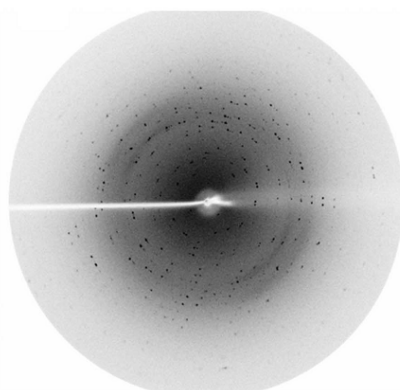
The crystal structures of the AVCP complexed with dioxane, piperazine and picolinic acid were studied. The structure of these three molecules is shown in Figure 3.4.1.



**Figure 3.4.1:** The structure of the compounds dioxane, piperazine and picolinic acid that was used for the study of the AVCP-complex structure.

### 3.4.1 Crystal structure of the dioxane bound AVCP

The structure of AVCP complexed with dioxane was refined to 1.98 Å resolution with R<sub>cryst</sub> value of 17.9 % and R<sub>free</sub> value of 26.7 % (Table 3.1). Similar to the native AVCP, the complex structure also belongs to the space group C2 with cell dimensions a = 79.6 Å, b = 35.2 Å, c = 49.5 Å. The Matthews coefficient is 2.0 Å<sup>3</sup> Da<sup>-1</sup> and the solvent content is 38.5 %. A single monomer is present in an asymmetric unit (Table 3.1). The N-terminal six residues are disordered and thus not visible in the electron density map of the complex structure similar to the native AVCP crystal structure.



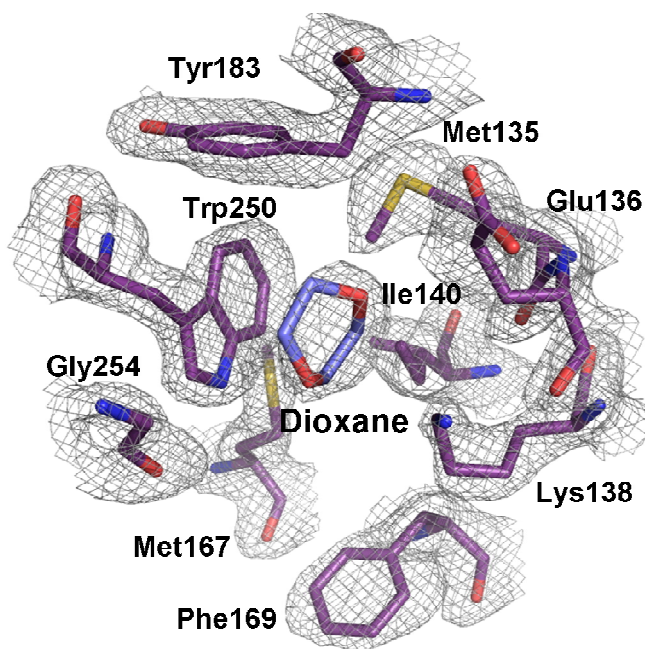
**Figure 3.4.1.1:** AVCP–dioxane complex crystal diffraction; the resolution at the edge of the plate is 1.98 Å.

The overall structure of the complex form is very similar to the native unbound form. The apo form of AVCP crystal structure reveals the presence of hydrophobic pocket composed of conserved residues on the surface of CP, which has also been reported in both the crystal structures of SCP and SFCP. Interestingly, a solvent-derived dioxane molecule was found to be bound to this hydrophobic pocket in the crystal structure of SCP (Lee *et al.*, 1998), suggesting that dioxane derivatives may prevent capsid–E2 hydrophobic interactions with the capsid hydrophobic pocket



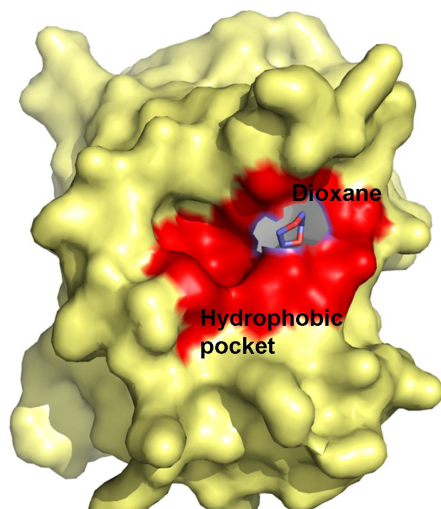
(Figure 2.4.5.3). As expected, dioxane-derived molecules showed antiviral properties against Sindbis virus (Kim et al, 2007; 2005).

AVCP is the third crystal structure of alphavirus CP, with SCP and SFCP; that confirms the presence of a structurally conserved hydrophobic pocket on the CP surface. Therefore, dioxane and similar molecules are expected to bind the hydrophobic pocket and inhibit virus budding by disrupting capsid–E2 interaction in all alphaviruses including Aura virus. Therefore, we determined the structure of AVCP in complex with dioxane. The electron density for the dioxane molecule is clearly visible in the difference Fourier map in the hydrophobic pocket of AVCP (Figure 3.4.1.2). The dioxane complex structure described in this report is the second structure of dioxane-CP complex among alphaviruses after SCP (PDB ID: 1WYK) (Lee *et al.*, 1998).



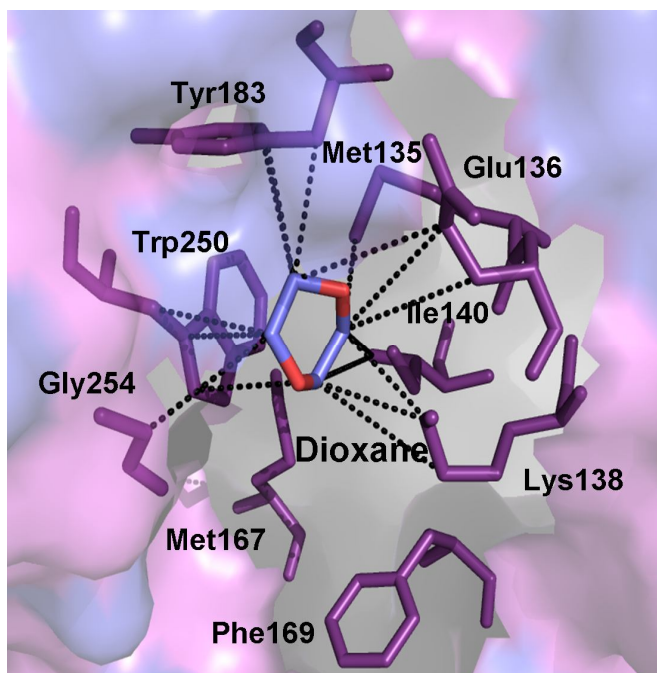
**Figure 3.4.1.2: AVCP-dioxane complex structure showing the dioxane binding pocket.** The hydrophobic pocket showing a bound dioxane molecule on the surface of AVCP crystal structure with 2Fobs-Fcalc electron density contoured at 1σ level. The residues present in the hydrophobic pocket are shown in violet and the dioxane molecule in blue. The electron density of the pocket is shown in gray.

The surface view of the AVCP with bound dioxane is shown in Figure 3.4.1.3. The hydrophobic pocket consists of two depressions one of which is occupied with a dioxane molecule. The position of this dioxane molecule is similar to that of the SCP crystal structure. However, the two dioxane molecules mimicking the binding of either Leu-X-Leu or Tyr-X-Leu are not found.



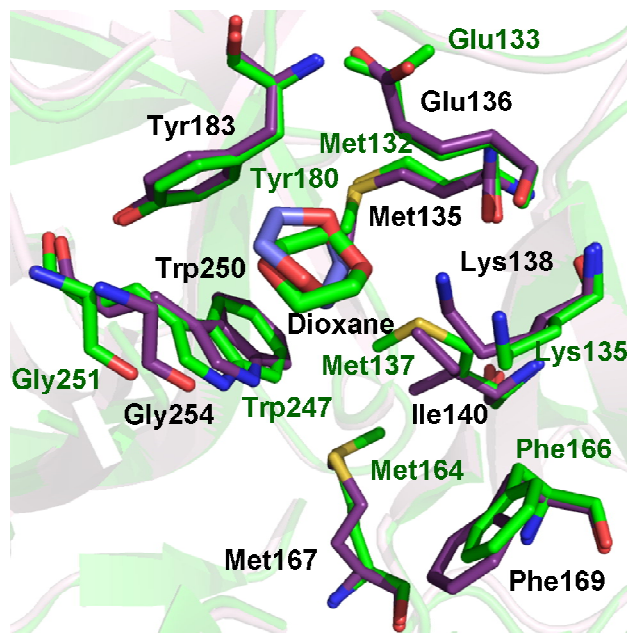
**Figure 3.4.1.3:** Surface view of the dioxane (blue) bound AVCP (yellow in color). The hydrophobic pocket residues are shown in red color.

In SCP, four residues (i.e. Met132, Glu133, Tyr180 and Trp247) contribute to major hydrophobic interaction with dioxane. However, in AVCP, seven residues (Met135, Glu136, Lys138, Ile140, Tyr183, Trp250 and Gly254) are in hydrophobic contact with the bound dioxane molecule (Figure 3.4.1.4).



**Figure 3.4.1.4:** The hydrophobic pocket of the AVCP complexed with dioxane as revealed by crystal structural studies. The interacting residues (violet) and dioxane molecule (blue) are shown as sticks. The hydrophobic interactions shown in dashed lines were analyzed by ligand explorer of RCSB PDB ([ligpro.sdsc.edu](http://ligpro.sdsc.edu)).

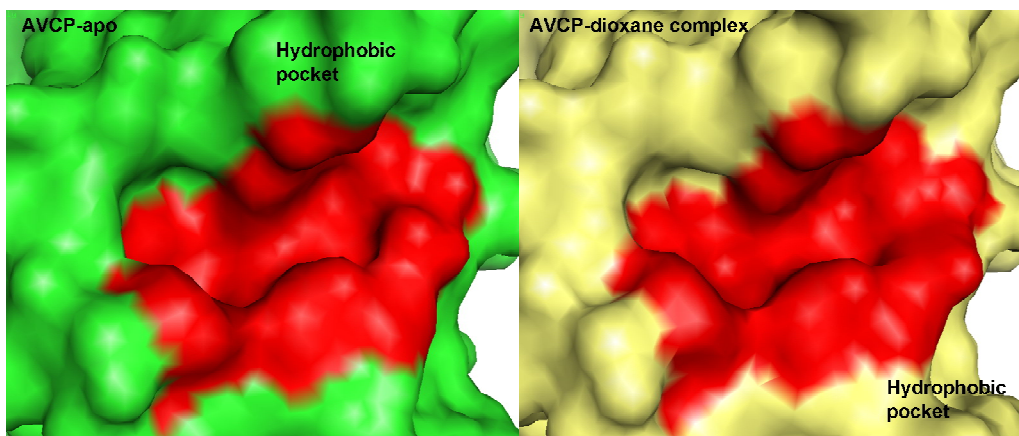
Dioxane binds to the hydrophobic pocket in the AVCP-dioxane complex in almost the same orientation as that of the SCP-dioxane structure with similar hydrophobic interactions. In the dioxane binding hydrophobic pocket, AVCP and SCP share highly conserved residues. Eight residues out of nine in this hydrophobic pocket are strictly conserved whereas Met137 of SCP is substituted with Ile140 in AVCP (Figure 3.4.1.5). Though SCP possesses the same number of conserved residues (except Met137), Lys135 and Gly251 are positioned slightly away from dioxane, rendering poor hydrophobic interactions with it (Figure 3.4.1.5). Thus, binding of dioxane to the CP hydrophobic pocket can be considered as characteristic for all the alphaviruses.



**Figure 3.4.1.5:** Superimposition of dioxane bound forms of AVCP (violet) and SCP (1WYK) (green) crystal structures highlighting the hydrophobic pocket residues.

### 3.4.2 Comparison of the dioxane bound complex structure with the apo form of AVCP

The superposition of AVCP-apo and AVCP-dioxane structures showed RMSD of 0.133 Å and dioxane binding does not affect the overall structure of AVCP. The comparison of the hydrophobic pocket in both the forms also does not show any significant deviation. The surface view of both the forms is shown in Figure 3.4.2.1. The minor deviation in the pocket is due to the side chain flexibility of Glu136 in between the apo and complex forms. Due to this minor change, the hydrophobic pocket in complex structure seems to be slightly wider as compared to the apo form of AVCP.

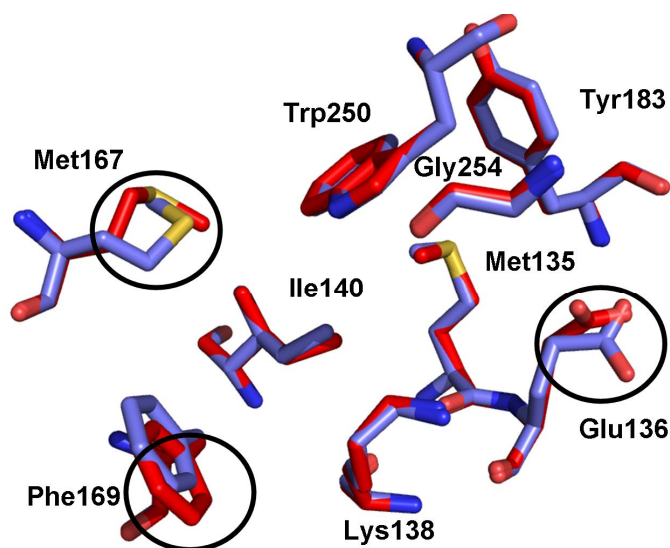


**Figure 3.4.2.1:** Structural comparison of the hydrophobic pocket in both apo (green in color) and dioxane bound complex form (yellow in color) of AVCP. The hydrophobic pocket forming residues are shown as red surface.

In order to analyze the conformational changes in the hydrophobic pocket between apo and dioxane bound AVCP, both the structures were aligned (Figure 3.4.2.2). The residues present in the hydrophobic pocket are Met135, Glu136, Lys138, Ile140, Met167, Phe169, Tyr183, Trp250 and Gly254. All the residues are static in position except the side chains of Glu136, Met167 and Phe169. The side chains of these three residues are flexible and adopt different conformations. Glu136 side chain just flips to the opposite direction in the complex form. However, these are minor variations and most probably do not effect the binding of the hydrophobic pocket towards different ligands.

The N-terminal arm or the dioxane molecule in hydrophobic pocket was present in the crystal structure of SCP. No structure with empty pocket is known yet for SCP. However, in SFCP, the N-terminal arm does not bind to the hydrophobic pocket due to some crystal arrangements. The SFCP crystal form was found to be different from the SCP ones in having the different orientation of Tyr180 and Trp247 (AVCP residues Tyr183 and Trp250). The different orientation of these two residues was supposed to the result of change in the conformation of the hydrophobic pocket. Thus, this might be responsible for the switch of hydrophobic pocket to function for capsid-capsid interaction towards capsid-glycoprotein interaction in SFCP. However, it is not clear whether the conformational changes are due to switch in the hydrophobic pocket function or it is just the property of different crystal structure form of other member of the alphavirus genus (Lee *et al.*, 1996). The SCP with truncated N-terminal arm was crystallized to get the exact information, but the dioxane molecule was found in the hydrophobic pocket. Thus, the structure with empty pocket for SCP is not available to compare the hydrophobic pocket.

The AVCP crystal structure of the apo form has an empty pocket, so it was compared with the hydrophobic pocket of dioxane bound form. The two residues Tyr183 and Trp250 were found exactly at the same position without any rearrangements in both the crystal forms. Thus, we can conclude that the hydrophobic pocket forming residues in the empty and the bound hydrophobic pocket show similar conformation. And just the two residues flipped in SFCP crystal structure might not responsible for the functional switching of the hydrophobic pocket.



**Figure 3.4.2.2:** The structural alignment of the hydrophobic pocket forming residues from apo (red) and dioxane bound form (blue) of AVCP. The circles represent the side chain flexibility of the residues.

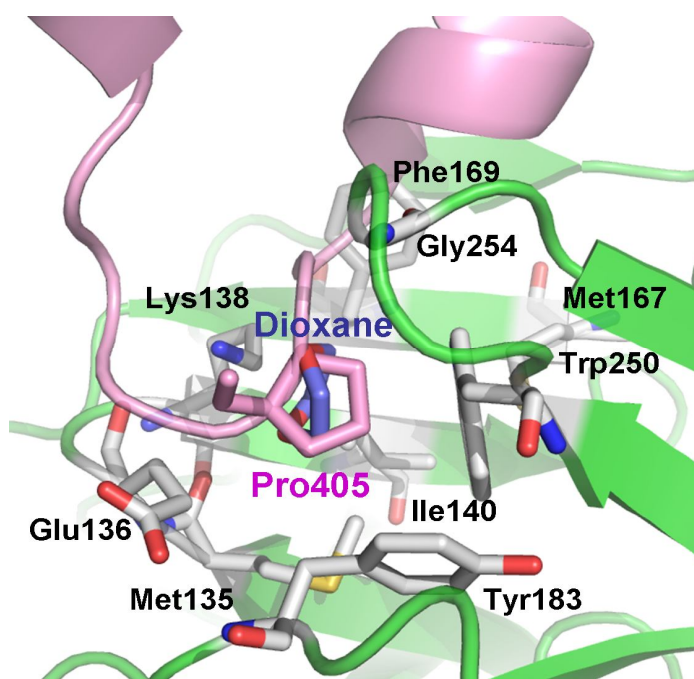
### 3.4.3 Inhibition strategy to disrupt capsid and glycoprotein interaction

The dioxane molecule bound to the hydrophobic pocket has been considered as the inhibitor of alphavirus CP (Lee *et al.*, 1998). A number of dioxane derivatives were designed by docking analysis for their binding to the hydrophobic pocket, which showed inhibitory activity against alphavirus infection. It is supposed that the inhibition is due to the binding of the dioxane molecule to the hydrophobic pocket for disrupting the interaction of E2 cytoplasmic tail with the CP. The capsid-E2 interaction is essential for the virus budding. Hence, by blocking the hydrophobic pocket for the E2 tail, dioxane molecule inhibits virus budding and thus acts as an inhibitor for the alphavirus infection. However, the detailed studies at molecular level are lacking. The exact mechanism for the dioxane to act as an inhibitor is still a mystery.

The interaction analysis of the E2 cytoplasmic tail with the CP is described in Chapter 2. The crystal structure of AVCP and the molecular models of E1 and E2 tails were fitted into the cryo-EM density map of VEEV. The Pro405 residue is found in the same hydrophobic pocket

through which the interaction takes place. The dioxane bound structure of AVCP was also fitted to the same density map and the analysis of the hydrophobic pocket was performed.

The interaction analysis of cdE2 and capsid reveals that dioxane occupies the same position as would be occupied by the pyrrolidine ring of Pro405 of cdE2 (Figure 3.4.3.1). Thus, competitive inhibition could be a plausible mechanism by which dioxane and its derivatives may disrupt capsid-E2 interaction in alphaviruses including Aura virus by obstructing the hydrophobic interaction of conserved Pro405 with CP. This assumption is based on the dioxane bound AVCP structural data where dioxane occupies the hydrophobic pocket on CP and structurally mimics the hydrophobic pyrrolidine ring of Pro405 in loop region of E2. Additionally, structural evidence shows binding of a dioxane molecule to the equivalent hydrophobic pocket on SCP and the inhibition of Sindbis virus infection by dioxane-based inhibitors supports this evidence (Kim *et al.*, 2007; 2005; Lee *et al.*, 1998).



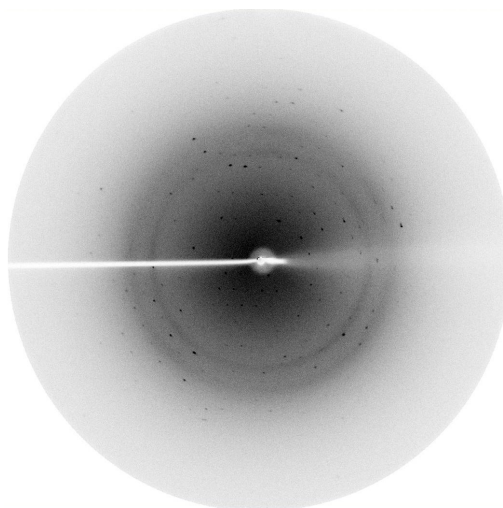
**Figure 3.4.3.1:** The dioxane molecule (blue) in the crystal structure of AVCP-dioxane complex occupies the same position as Pro405 (pink) present in the helix-loop-helix motif of cdE2. A homology model of cdE2 was combined with the AVCP crystal structure. The residues in AVCP that are involved in the interaction are shown as gray sticks.

Figure 3.4.3.1 shows the overlapping of the dioxane molecule and Pro405 residue of the E2 tail. Earlier studies suggested that dioxane might be involved in preventing the interaction of the Tyr400 of SINV E2 tail to the hydrophobic pocket of capsid (Lee *et al.*, 1998). However, the cryo-EM density fit analysis described in Chapter 2 shows that the corresponding Tyr401 residue of

Aura virus E2 tail orients away from the hydrophobic pocket of CP and does not reach near the pocket for binding. Also, the structure of the dioxane mimics the pyrrolidine ring of proline residue and provides support to our analysis for the direct binding of Pro405 of E2 tail to the hydrophobic pocket of CP. However, it is further essential to perform the interaction studies through mutational and structural analysis to confirm this binding strategy.

#### 3.4.4 Complex of AVCP with piperazine

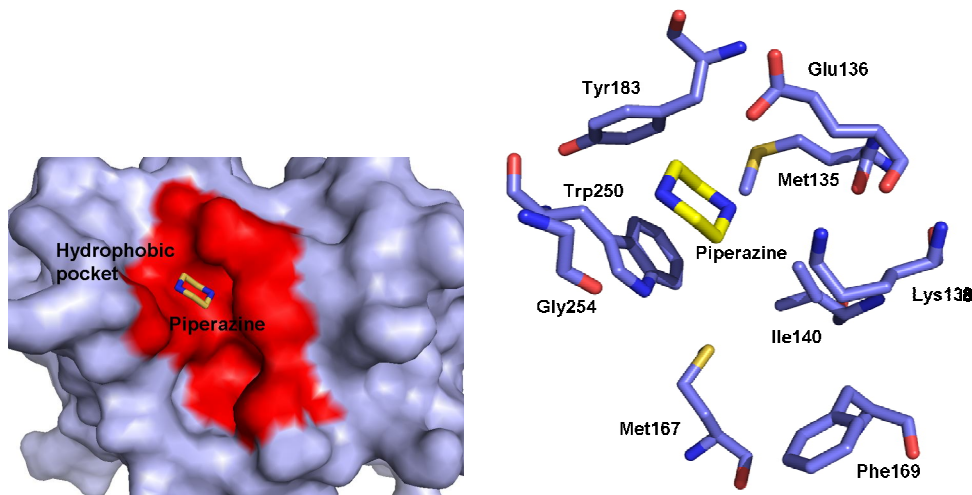
As the dioxane molecule occupies the hydrophobic pocket of AVCP and disrupts the CP-glycoprotein interaction, other molecules similar to dioxane were designed and co-crystallization trials were performed. The crystal of the native AVCP was soaked in piperazine to get the structure of AVCP-piperazine complex. The crystal was diffracted up to 2.2 Å resolution and the data were collected at home source (Figure 3.4.4.1). After refinement, the final structure was analyzed for the binding of piperazine to the AVCP. The Matthews coefficient and the solvent contents are very similar to the other AVCP structure forms and found to be  $1.95 \text{ \AA}^3 \text{ Da}^{-1}$  and 37 % respectively. There is one molecule present per asymmetric unit. The Rvalue and Rfree for the piperazine bound AVCP are 19.8 and 27.6 % respectively. The data collection and refinement statistics are given in Table 3.1.



**Figure 3.4.4.1:** The diffraction pattern of AVCP-piperazine complex crystal; the resolution at the edge of the plate is 2.2 Å.

The structure analysis of the piperazine-bound AVCP shows the presence of the piperazine molecule in the same hydrophobic pocket. The piperazine molecule binds to the one groove of the hydrophobic pocket which is lined by the hydrophobic residues. The hydrophobic pocket residues show ionic as well as hydrophobic interactions with the piperazine. The binding is very similar to

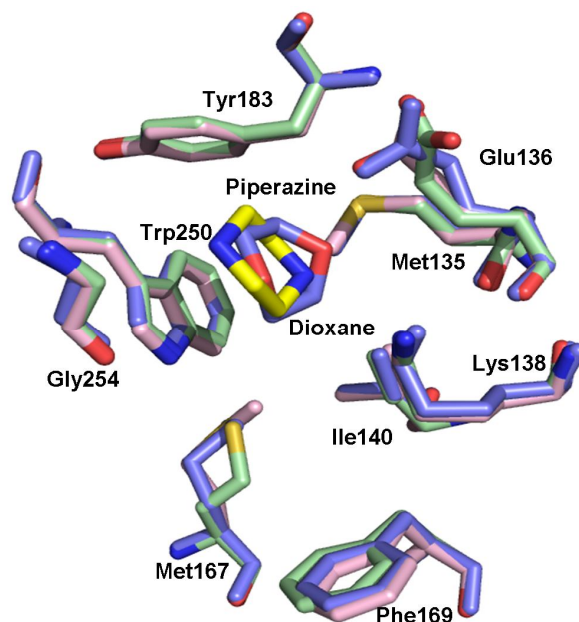
the binding of the dioxane molecule in the pocket. Thus, it can be inferred that piperazine can also play role as an inhibitor against alphavirus budding process (Figure 3.4.4.2).



**Figure 3.4.4.2:** The surface view of the AVCP in complex with the piperazine (left panel). The piperazine molecule (yellow sticks) is present in the hydrophobic pocket. The hydrophobic pocket residues are shown in red color. The right panel shows the piperazine molecule surrounded by the hydrophobic pocket residues. The residues are shown as sticks.

The AVCP-piperazine bound structure was compared with the apo-AVCP and AVCP-dioxane bound structure (Figure 3.4.4.3). The rms deviation of the piperazine bound form with the apo and dioxane bound forms was calculated and found to be 0.147 and 0.178 respectively. The binding pattern of the piperazine is almost similar to that of dioxane molecule. However, as can be seen from the figure, there is some displacement in the position of these two molecules. The hydrophobic residues surrounding the ligand molecules also show same conformations except the side chains of Glu136 and Met167 where the variation is found in the dioxane bound AVCP. The apo and the piperazine bound forms show very similar conformation for the hydrophobic pocket residues.



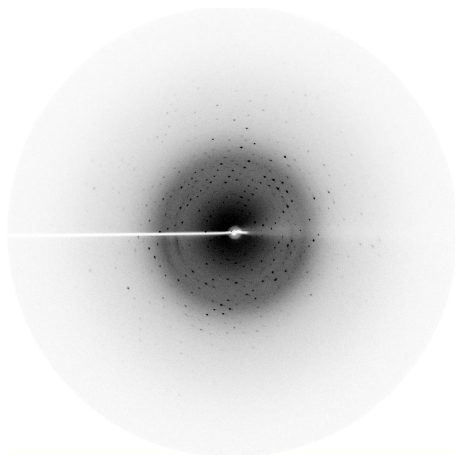


**Figure 3.4.4.3:** Comparison of the apo (pink), dioxane bound (green) and piperazine bound (blue) AVCP. The overall binding pattern is similar to the dioxane bound form with some side chain modifications.

The results show the molecules having similar structure as that of dioxane (which can properly fit into the hydrophobic pocket) can be used as inhibitors of alphavirus budding process. Not only dioxane and its derivatives can act as inhibitor molecules against alphavirus infection, however, piperazine and its derivatives also can be tested for the inhibition of alphavirus infection. Thus, the possibilities of designing new inhibitor molecules have increased which can compete with the binding of E2 tail to the hydrophobic pocket of CP.

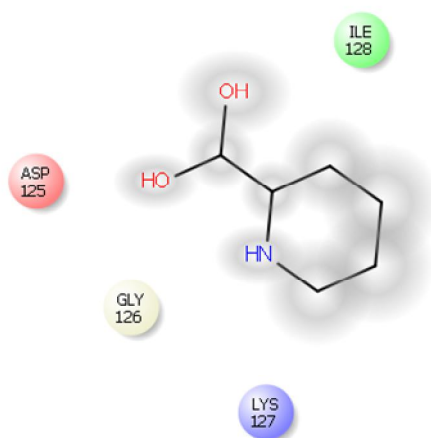
### 3.4.5 Complex of AVCP with picolinic acid

In the series of complex formation of AVCP with potential antiviral compounds, picolinic acid was used for the complex preparation. The high resolution data of 1.73 Å were collected at home source and the structure determination and refinement were performed (Figure 3.4.5.1). The Matthews coefficient and the solvent content were almost same as that for the dioxane bound structure ( $2.03 \text{ \AA}^3 \text{ Da}^{-1}$  and 39.5 % respectively). The Rfactor and Rfree are 18.0 and 23.5 % respectively for the picolinic acid bound form of the AVCP.



**Figure 3.4.5.1:** The diffraction pattern of AVCP-picolinic acid complex crystal; the resolution at the edge of the plate is 1.73 Å.

The refinement statistics are given in Table 3.1. The interaction analysis was performed and the ligand interaction figure was prepared using Maestro ligand interaction wizard with a cut off value of 5 Å (Figure 3.4.5.2). The figure shows the interaction of the picolinic acid with Asp125, Gly126, Lys127 and Ile128 residues of the CP.



**Figure 3.4.5.2:** The interaction of the picolinic acid at the surface of AVCP. The figure was prepared in the ligand interaction wizard of Maestro.

The picolinic acid binds to the surface of the capsid and interacts with the neighboring symmetry molecule. It is surrounded by the other symmetry related molecules. The surface view of the AVCP with its symmetry molecules is shown in Figure 3.4.5.3a. The picolinic acid is found to be hidden or protected from the environment by the capsid molecules that surround it.

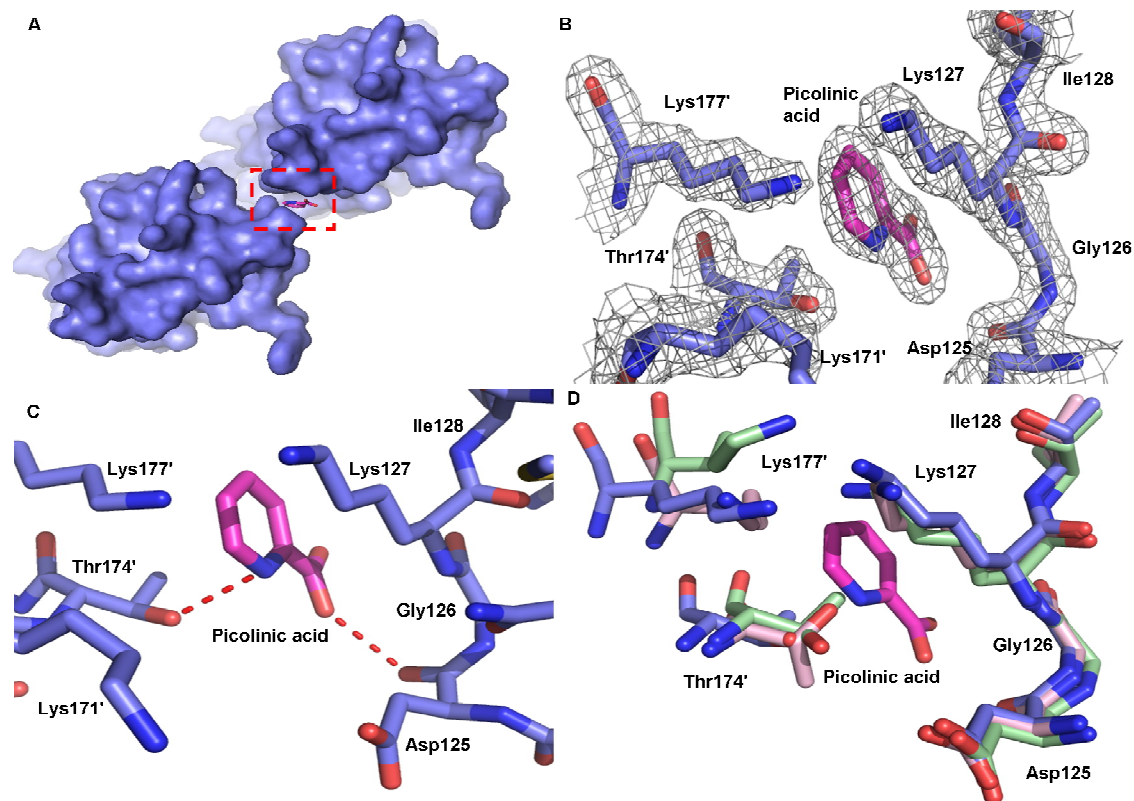
The structure of the picolinic acid-complex shows the empty hydrophobic pocket with no density for the compound. The electron density for the picolinic acid was missing in the

hydrophobic pocket. However, further analysis shows the presence of clear density of the picolinic acid in some other pocket at the surface of AVCP at 1.0 sigma level (Figure 3.4.5.3b). Hence, this was the interesting finding that picolinic acid binds to some other place instead of the hydrophobic pocket.

Structural analysis of the picolinic acid bound complex demonstrates the interaction of the picolinic acid with the Asp125, Gly126, Lys127 and Ile128 residues of AVCP. The O atom of the picolinic acid forms the ionic interaction with the Asp125 residue main chain. However, other residues are involved in the hydrophobic interactions. On the other side, the ligand forms the contacts with other neighboring symmetry related molecule through residues Lys171, Thr174 and Lys177. The N atom in the ring of the picolinic acid shows H-bonding with the Thr174 residue of the symmetry molecule (Figure 3.4.5.3c).

The structural comparison of the picolinic acid bound AVCP with the apo and dioxane bound forms of AVCP is shown in Figure 3.4.5.3d. The RMSD value of the structure with the native AVCP is 0.260 while with the dioxane bound AVCP, it is 0.318. As the picolinic acid binds to the surface and does not adopt any conformational changes to the hydrophobic pocket, it shows more deviation with the dioxane bound form in comparison to the native AVCP. The surface residues which show interaction with the picolinic acid are compared and the significant variation in the symmetry related molecules of the different ligand bound AVCP is observed. Lys177' shows the maximum conformational variation in between different forms of AVCP.

Thus, the picolinic acid binds at the surface of the AVCP and shows significant variations with the other crystal forms of AVCP for the surface residues. The dioxane bound form shows the most structural variation. The picolinic acid binding to the surface of AVCP still needs more detailed studies to understand if the binding of the picolinic acid inhibits the alphavirus infection or any other role of its binding.



**Figure 3.4.5.3: The interaction strategy of picolinic acid to AVCP.** **a:** The surface view of the AVCP with its symmetry related molecule. The picolinic acid is bound to the surface of AVCP. The picolinic acid resides in between the AVCP and its symmetry related molecule. **b:** The electron density of the picolinic acid is clearly visible at the surface of AVCP between two neighboring symmetry molecules. The density map was formed at the sigma cut off of 1.0. **c:** The interaction of the picolinic acid with the AVCP and its symmetry related molecule. Asp125 forms the ionic interaction with the -OH group of the picolinic acid, however its N atom present in the ring structure forms the interaction with Thr174 of the neighboring symmetry molecule. **d:** Structural comparison of the AVCP-picolinic acid bound structure (blue) with the apo (pink) and dioxane bound (green) forms of AVCP.

**Table 3.1:** Data collection and refinement statistics for AVCP in complex with different ligands.

	AVCP-dioxane Complex	AVCP-piperazine complex	AVCP-picolinic acid Complex
Crystallographic Data			
Space group	<i>C2</i>	<i>C2</i>	<i>C2</i>
Cell dimensions (Å)	$a=79.6$ , $b=35.2$ , $c=49.5$	$a=79.256$ , $b=34.839$ , $c=49.169$	$a=79.084$ , $b=36.211$ , $c=49.330$
Resolution range (Å)	50.0 – 2.21	50.0 – 2.25	50.0 – 1.73

(Last Shell)	(2.02–1.98)	(2.25–2.21)	(1.76–1.73)
Completeness (%) (Last Shell)	86.2 (45.7)	79.6 (41.6)	94.2 (56.9)
$R_{\text{merge}}^{\dagger}$ (%) (Last Shell)	4.8 (28.2)	6.3 (25.7)	3.0 (32.9)
Mean $I/\sigma(I)$ (Last Shell)	20.3 (1.8)	14.8 (2.00)	30.22 (1.99)
No. of observed reflections	24420	4614	12938
No. of unique reflections (Last Shell)	7989 (160)	4847 (21)	13625 (406)
Molecules per asymmetric unit	1	1	1
Matthews coefficient ( $\text{\AA}^3 \text{Da}^{-1}$ )	1.99	1.95	2.03
Solvent content (%)	38.5	37.07	39.52
Multiplicity (Last Shell)	3.1 (1.9)	2.7 (1.7)	3.4 (2.3)
Refinement			
No. of Residues	152	152	155
Water molecule	102	76	132
$R_{\text{cryst}}$ (%)	17.9	19.8	18.0
$R_{\text{free}}$ (%)	26.75	27.6	23.5
Average $B$ -factor ( $\text{\AA}^2$ )	31.43	32.58	33.33
r.m.s.d on bond lengths ( $\text{\AA}$ )	0.015	0.004	0.016
r.m.s.d on bond angles ( $\text{\AA}$ )	1.926	0.947	1.888
Ramachandran plot (%)			
Preferred	95.33	94.00	93.84
Allowed	4.0	4.67	4.79
Outliers	0.67	1.33	1.37

$\dagger R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - [I(hkl)]|}{\sum_{hkl} \sum_i I_i(hkl)}$ , where  $I_i(hkl)$  is the  $i$ th observation of reflection  $hkl$  and  $[I(hkl)]$  is the weighted average intensity for all observations  $i$  of reflection  $hkl$ .

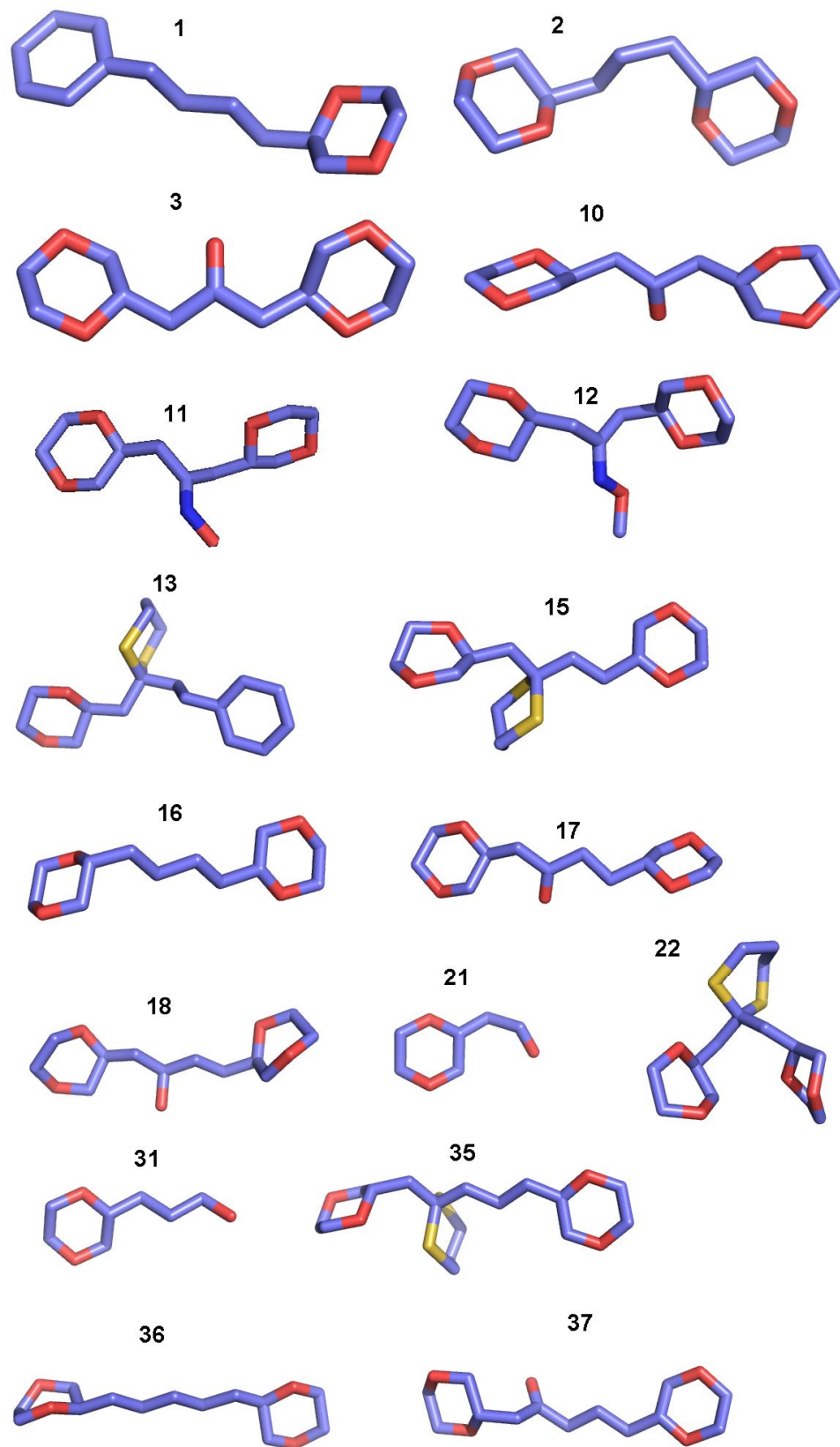
### **3.4.6 Docking analysis**

The dioxane molecule bound to the hydrophobic pocket provided a hint for the discovery of dioxane based antivirals to inhibit virus budding and virus infection. Two dioxane molecules were present in one Chain of SCP-dioxane complex which were present in two separate grooves of the hydrophobic pocket. Hence, the compounds with two ring structures were supposed to bind the CP hydrophobic pocket more efficiently. Therefore, to get the proper knowledge of binding, different compounds were used for the docking analysis with the AVCP.

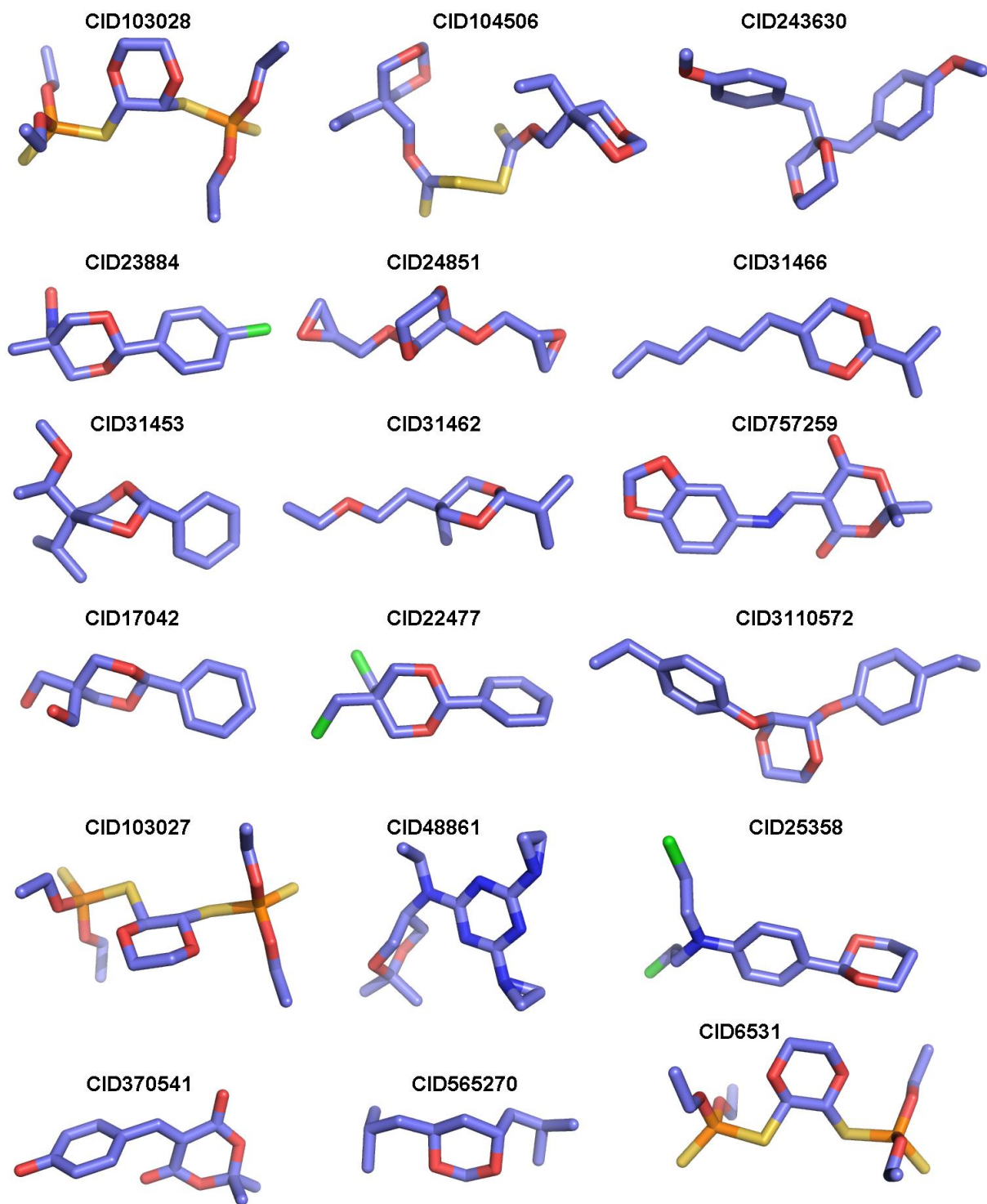
#### **3.4.6.1 Docking of dioxane based and similar compounds**

A number of compounds containing two dioxane rings have been designed and used for the inhibition studies of the SINV (Kim *et al.*, 2007; 2005). In the present study, we have analyzed those structures through computational docking analysis. The molecular docking was performed using the compounds and compared with the inhibition results with previous viral infection studies. All the compounds were docked and the compounds that show good docking results are shown in Figure 3.4.6.1.1. The compounds were drawn in Chemdraw and prepared using the ligprep wizard of Maestro. The docking analysis was performed in Glide which generated a number of poses. All the poses were visualized in PyMol. The best pose was selected and Glide score was also considered in selecting the docked structure. The compounds possess two ring structures connected through either the aliphatic chain or aromatic ring structure. The compounds which did not efficiently show virus inhibition were also used in docking analysis to find out if they can get docked well inside the hydrophobic pocket. Thus, the docking studies can be performed for virtual screening of the compound library against alphavirus budding.

In order to screen compounds by computational approach, several other potential inhibitory molecules were derived from PubChem and the molecular docking was performed using Glide. The selected compounds that show good Glide score and best pose were chosen and shown in Figure 3.4.6.1.2. The docked structures are selected after visually investigating the complex structure. The compounds contain two ring structures connected to each other through aliphatic carbon chain having different modifications.



**Figure 3.4.6.1.1:** Compounds used for the docking analysis based on previous studies. The numbering written above the compound structure is same as given in the inhibition studies (Kim *et al.*, 2007; 2005)

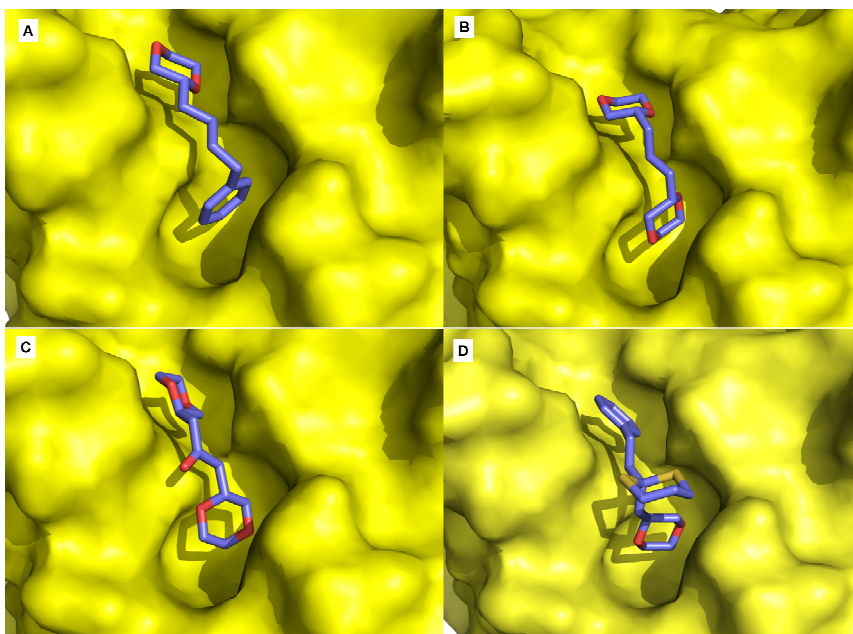


**Figure. 3.4.6.1.2:** Compounds used for docking analysis which were derived from PubChem based on the structure of dioxane. The PubChem ID is written above the structure of all the compounds.



### 3.4.6.2 Binding of the compounds to the hydrophobic pocket

The docking was performed for all the compounds described above. As expected, the two rings of the compounds are fitted into the two separate grooves of the hydrophobic pocket as described for one of the Chain of SCP-dioxane complex. Figure 3.4.6.2.1 shows four compounds docked into the hydrophobic pocket. The binding to the pocket is very similar to each other for all the four compounds. Other compounds were also showing almost similar conformation. The binding to two different grooves of the hydrophobic pocket is common to all the compounds. Thus, it was supposed that instead of a single dioxane molecule, the two ring structure would bind more efficiently to the hydrophobic pocket and inhibit more potently the binding of the E2 cytoplasmic tail.

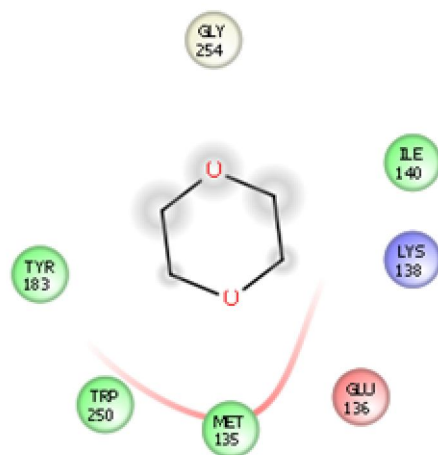


**Figure 3.4.6.2.1:** The binding of four different compounds (1, 2, 3, 13) into the hydrophobic pocket of the AVCP. All the four compounds are showing similar binding pattern by binding to both the grooves of the hydrophobic pocket through the aromatic rings.

### 3.4.6.3 Interaction of the dioxane and other compounds to the hydrophobic pocket

The interaction of the dioxane to the hydrophobic pocket has been described in the previous sections. This section includes the interaction comparison of the single dioxane molecule and similar compounds containing two dioxane moieties or two ringed structures. The interaction was studied in Maestro ligand interaction wizard and the figures were prepared in its ligand interaction wizard. The single dioxane molecule shows the interaction with the hydrophobic pocket through interacting with the residues present in one groove of the pocket. The interaction

takes place with Met135, Glu136, Lys138, Ile140, Tyr183, Trp250 and Gly254 residues of the hydrophobic pocket (Figure 3.4.6.3.1).

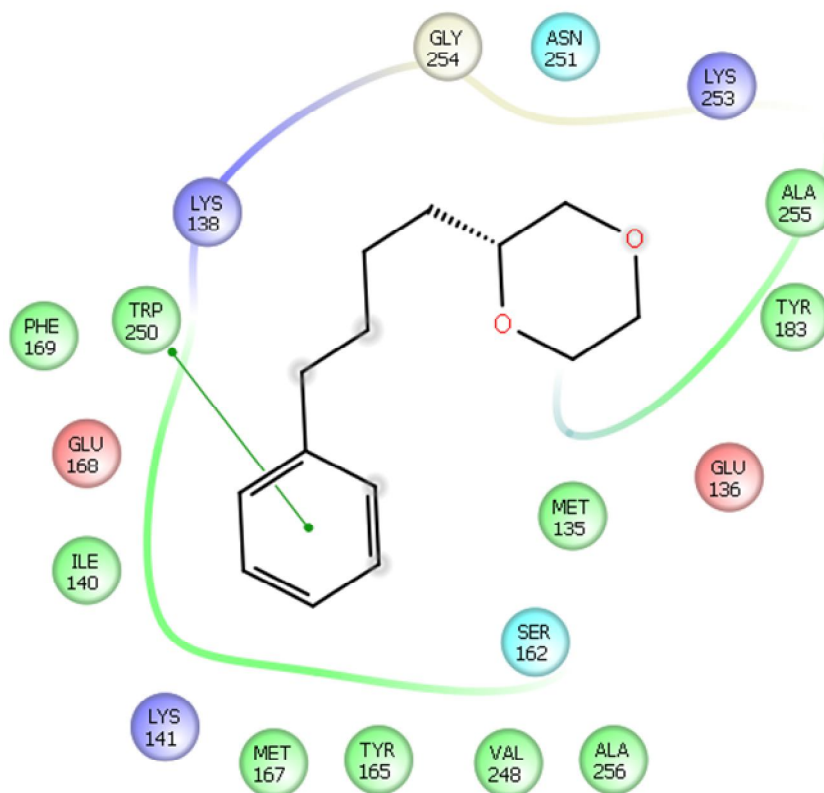


**Figure 3.4.6.3.1:** The dioxane molecule present in the hydrophobic pocket shows the hydrophobic interactions with the pocket residues. The figure was prepared in the ligand interaction wizard of Maestro.

The docking results for the compounds having two ringed structure show more efficient binding to the pocket, as the dissociation of the two rings will be difficult as compared to the single ring structure. So, in the absence of the dissociation of the ligand, the E2 glycoprotein will not be able to further bind to the hydrophobic pocket which will prevent the budding of the alphavirus. Thus, the two ring compounds can inhibit the virus infection more efficiently.

The two ring compound shows more interactions to the hydrophobic pocket as it includes the residues from both the grooves of the pocket. It reveals half of the interaction similar to the dioxane molecule, while the half other interactions are not present for the dioxane molecule. The interacting residues include Met135, Glu136, Lys138, Ile140, Lys141, Ser162, Tyr165, Met167, Glu168, Phe169, Tyr183, Val248, Trp250, Asn251, Lys253, Gly254, Ala255 and Ala256. The hydrophobic interaction dominates in both the cases. The interactions are shown to a cut off value of 5 Å. Within this distance all the interacting residues are considered. All the compounds show similar interactions with the hydrophobic pocket as described for the compound 1 in Figure 3.4.6.3.2. However, one interaction is present that represent the  $\pi$ - $\pi$  interaction between the two aromatic rings of the Trp250 and the ring of the compound. This interaction is absent in those compounds in which both the rings are having dioxane moieties. Some of the compounds are having the ring structure or the branch structure in the connector of the two ring structures. These ring or the branch structures are pointed outwards and do not influence the binding of the two rings of compound. These structures do not show any interaction with the hydrophobic pocket residues.

Thus, these modified structures are oriented outside the pocket and projected outwards at the surface of the AVCP. Hence, it can be concluded that all the compounds follow a common interaction strategy to bind to the hydrophobic pocket. The modification in the compound structure leads to minor changes in the interaction pattern.



**Figure 3.4.6.3.2:** The interaction of the compound 1 with the hydrophobic pocket residues of the AVCP is shown. Maestro's ligand interaction wizard is used for the preparation of the figure. The cut-off value for the interaction was set to 5 Å. The green line shows the  $\pi$ - $\pi$  interaction between the aromatic rings of Trp250 and one ring of the compound 1.

The Glide Score and the Glide Emodel for the compounds are shown in Table 3.2. However, the data in the table show the incomparable results for the docking analysis and the inhibition studies done previously. The compounds having the two ringed structure are showing good Dock Score as compared to the compounds having single ring (Compound 31). Thus, two dioxane molecules will prevent the binding of E2 tail to the hydrophobic pocket more efficiently according to structural studies.

**Table 3.2:** The docking studies with the AVCP for the dioxane analogs.

Comp ound	Compound name	Glide Score	Glide Emodel
1	(R)-2-(4-phenylbutyl)-1,4-dioxane	- 4.372	- 30.284
2	(S)-2-(3-((R)-1,4-dioxane-2-yl)propyl)-1,4-dioxane	- 3.865	- 23.536
3	1-((R)-1,4-dioxan-2-yl)-3-((S)-1,4-dioxan-2-yl)propan-2-one	- 4.406	- 25.949
10	1,3-Bis{(R)-[1,4]dioxan-2-yl}propan-2-ol	- 3.846	- 24.366
11	1,3-Bis{(R)-[1,4]dioxan-2-yl}-2-oximinopropane	- 3.139	- 23.427
12	1,3-Bis{(R)-[1,4]dioxane-2-yl}-2-(methyloximino)propane	- 4.163	- 25.212
13	(R)-2-(2-Phenethyl-[1,3]dithian-2-ylmethyl)-[1,4]dioxane	- 3.321	- 29.863
15	1,4-Bis{(R)-[1,4]dioxan-2-yl}-2-([1,3]dithian-2-yl)butane	- 3.292	- 26.837
16	1,4-Bis{(R)-[1,4]dioxan-2-yl}butane	- 3.900	- 24.926
17	1,4-Bis{(R)-[1,4]dioxan-2-yl}-2-butanone	- 4.841	- 33.092
18	1,4-Bis{(R)-[1,4]dioxan-2-yl}butan-2-ol	- 3.842	- 27.059
21	(S)-[1,4]Dioxan-2-yl-acetaldehyde	- 3.770	- 20.264
22	1,3-Bis{(R,S)-[1,4]dioxan-2-yl}-2-[1,3]dithian-2-yl)propane	- 3.474	- 26.412
31	(R)-3-([1,4]Dioxan-2-yl)propan-1-ol	- 2.937	- 19.961
35	1,5-Bis{(R)-[1,4]dioxan-2-yl}-2-([1,3]dithian-2-yl)pentane	- 3.907	- 30.015
36	1,5-Bis{(R)-[1,4]dioxan-2-yl}pentane	- 4.121	- 29.822
37	1,5-Bis{(R)-[1,4]dioxan-2-yl}pentan-2-one	- 4.681	- 30.797

The docking analysis for the other compounds derived from PubChem was performed. Compound 22477 was found to show the highest Glide Score and 24851 shows the lowest Glide Score (Table 3.3). Some of these PubChem derived compounds show relatively less Glide Score as compared to the compounds that were previously used for the study.

Thus, the virtual screening of the compounds for binding to the hydrophobic pocket is highly useful in finding the inhibitors against the budding process of alphaviruses which leads to inhibit virus infection. Further, the compounds found to be potent in computational docking studies, can be tested for inhibition of the virus infection. However, it is possible that the compounds showing good docking scores do not exhibit antiviral activity. Also, the cytotoxicity of the compounds is also of the major concern along with the antiviral assays, because if a compound is cytotoxic, then it will kill the infected cells and the compound will be of no use.

**Table 3.3:** The Glide Score and Glide Emodel of the compounds derived from PubChem according to their rank for docking score.

Rank	Compound ID	Glide Score	Glide Emodel
1	CID22477	- 4.541	- 25.688
2	CID23884	- 4.385	- 28.555
3	CID6531	- 4.237	- 41.090
4	CID3110572	- 4.237	- 34.077
5	CID370541	- 3.971	- 30.485
6	CID103027	- 3.904	- 40.500
7	CID757259	- 3.870	- 34.535
8	CID243630	- 3.661	- 33.505
9	CID25358	- 3.565	- 32.338
10	CID17042	- 3.407	- 27.717
11	CID31453	- 3.4	- 23.851
12	CID103028	- 3.3	- 34.386
13	CID565270	- 3.034	- 21.097
14	CID48861	- 2.533	- 15.147
15	CID31462	- 2.4	- 20.275
16	CID104506	- 2.149	- 28.91
17	CID31466	- 1.74	- 12.088
18	CID24851	- 1.209	- 25.755

### 3.5 Conclusion

The crystal structure of AVCP in complex with dioxane showed a dioxane molecule bound to the hydrophobic pocket of CP. The similarity of the alphavirus capsid hydrophobic pocket and binding of dioxane to the same hydrophobic pocket in the crystal structure of CP from two alphaviruses (SCP and AVCP) signifies that a broad-spectrum inhibitor that targets CP-E2 interactions during virus budding can be developed. In addition, based on Pro405 interaction with CP, a structure-based approach can be employed for the design and development of novel pyrrolidine derived alphavirus-specific antivirals. The crystal structure of the AVCP-dioxane complex demonstrates the most striking feature of dioxane that exclusively occupies the position of Pro405 in the CP hydrophobic pocket. Thus, it is proposed that dioxane based-derivative

### Chapter 3: Assessment of the complexes of AVCP with dioxane and similar compounds

molecules would compete with the binding of Pro405 in the CP hydrophobic pocket and disrupt the CP-E2 interaction in Aura virus.

The crystal structures of the complexes of AVCP with other ligands (picolinic acid and piperazine) were also analyzed. The piperazine is found to be placed in the same way as dioxane binds to the hydrophobic pocket with similar interactions. The clear electron density is visible for the presence of piperazine to the hydrophobic pocket. However, the electron density of the picolinic acid in the hydrophobic pocket is not visible and it is present at the surface of the capsid. The picolinic acid shows interaction with the AVCP along with the neighboring symmetry molecules. These molecules surround the picolinic acid and seem to hide it from the surrounding environment. However, more detailed studies for the binding of picolinic acid are needed. It needs to evaluate the importance of the position where picolinic acid binds. Though, the studies regarding the complex structure of the CP with picolinic acid provides the clue for some interaction at the surface.

Different compounds were designed on the basis of previous inhibition studies (Kim *et al.*, 2007; 2005) and some are derived from PubChem. These compounds were docked into the AVCP for the virtual screening of the compounds. The antiviral activity of the compounds from previous studies was also compared with the docking results of the same compounds. The compounds having the two ring structures were found to be more efficiently bound to the capsid. These two different rings of the compounds bind to the two grooves present in the hydrophobic pocket. As the dissociation energy would be more for the compounds with two ring structure, it would be difficult to replace the compound by any other molecule like E2 cytoplasmic tail. Thus, the docking results show the good scores for the compounds having two rings; however, the antiviral activity for some of the double ring compounds was very less as compared to the single ring structures. Thus, we can say that some other factors also influence the antiviral activity of the compounds. Although, the docking results provide a hint for the compounds that binds to the pocket and those can act as a good inhibitor against the virus infection. Hence, these compounds might be useful in disrupting the CP-E2 interaction that leads to virus budding. Thus, dioxane and similar compounds would be highly useful for future inhibition studies against the alphavirus infection.

### 4.1 Abstract

In this Chapter, we report the crystal structure of the trans-active form of AVCP determined to 1.81 Å resolution. Structural comparisons of the active form with the crystal structures of available substrate-bound mutant and inactive blocked forms of the capsid protease identifies conformational changes in the active site, the oxyanion hole and the substrate specificity pocket residues, which could be critical for rational drug design.

The alphavirus capsid protein (CP) is a serine protease that possesses cis-proteolytic activity essential for its release from the nascent structural polyprotein. The released CP further participates in viral genome encapsidation, nucleocapsid core formation followed by its attachment to glycoproteins and virus budding. Thus, protease activity of the alphavirus capsid is a potential anti-alphaviral target to arrest capsid release, maturation and structural polyprotein processing. However, the discovery of capsid protease inhibitors has been hampered due to the absence of its crystal structure in the active form. The alphavirus capsid protease is an attractive antiviral therapeutic target.

Our structural studies unveil the structural features of the trans-active protease which has been previously proposed to exist in the natively unfolded form (Morillas *et al.*, 2008). The different enzymatic forms have been structurally compared to reveal conformational variations in the active and substrate binding site. The flexible active site residue Ser218, the disordered C-terminal residues after His261, and the presence of a water molecule in the oxyanion hole of AVCPΔ2 (AVCP with a deletion of the last two residues at the C terminus) reveal the effect of the C-terminal Trp267 deletion on the enzyme structure. New structural data reported in this study will be useful in substrate specificity characterization and structure-based development of antiviral drugs against the alphavirus capsid protease.

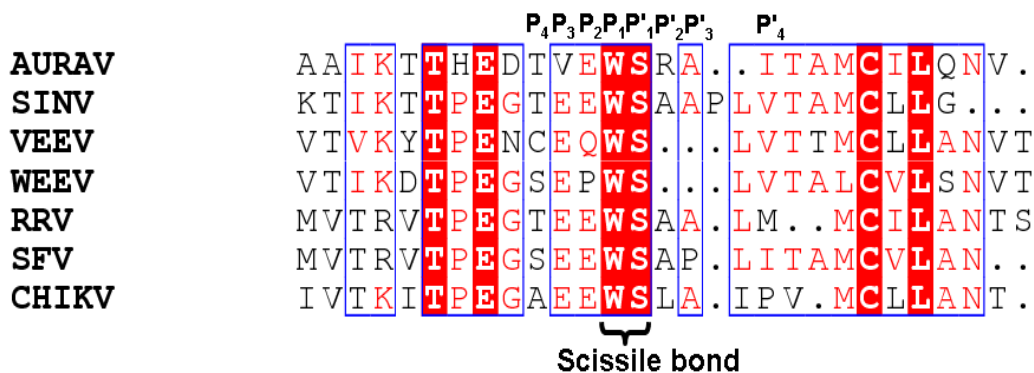
### 4.2 Introduction

The alphavirus infection in humans can cause fever, rash, encephalitis and polyarthrititis. Due to the epidemic outbreak of CHIKV in 2005-2006 in the Indian subcontinent, it is considered to be a re-emerging pathogen and a potential public health threat. Currently no antiviral drug or vaccine is available against alphaviruses. The alphavirus CP is present at the amino-terminus of the structural polyprotein and has been reported to be a chymotrypsin-like serine protease. The first step in structural polyprotein processing is the autocatalytic cleavage of CP to release itself from rest of the polyprotein (Choi *et al.*, 1991; Melancon & Garoff, 1987). This released CP

## Chapter 4: Structural insights into the active form of the alphavirus capsid protease

performs multiple functions in the virus life cycle including the formation of capsomers by intermolecular interactions with other CP monomers, encapsidation of the genomic RNA to form the nucleocapsid cores, and interaction with the cytoplasmic domain of glycoproteins that is essential for the virus budding process (Hong *et al.*, 2006; Zhao *et al.*, 1994; Metsikkö & Garoff, 1990; Weiss *et al.*, 1989; Hahn *et al.*, 1988; Vaux *et al.*, 1988).

The alphavirus CP consists of two major domains. The N-terminal domain is highly disordered, rich in basic amino acids, and functions during encapsidation of viral genomic RNA (Owen & Kuhn, 1996; Forsell *et al.*, 1995; Geigenmüller-Gnirke *et al.*, 1993; Weiss *et al.*, 1989). The C-terminal domain possesses cis-autoproteolytic activity and is inactivated after cleavage of the scissile bond between Trp-Ser residues (AVCP residues: W267-S268). The Trp-Ser residues containing the scissile bond are conserved among all alphaviruses and are present at the C-terminal end of CP (Figure 4.2.1). After cleavage, the free carboxylic group of conserved Trp267 at P<sub>1</sub> position interacts with the catalytic triad and remains bound in the S<sub>1</sub> pocket of the CP. This bound carboxyl terminus tryptophan residue in the active site inactivates the protease and blocks further trans-cleavage activity (Skoging & Liljestrom, 1998; Choi *et al.*, 1996; Tong *et al.*, 1993; Choi *et al.*, 1991). The active site molecular architecture and the catalytic triad are well conserved among all serine proteases including alphavirus CPs (AVCP: His144, Asp163 and Ser218) and exhibit similar spatial architecture in the active site (Choi *et al.*, 1991; Hahn & Strauss, 1990). Additionally, the GDSG motif containing the active site nucleophilic serine residue that is well conserved in chymotrypsin-like serine proteases is also found to be completely conserved in the CP of alphaviruses (AVCP: 215GDSG219) (Choi *et al.*, 1996; Tong *et al.*, 1993; Choi *et al.*, 1991).



**Figure 4.2.1: Multiple sequence alignment of the residues at scissile bond.** P<sub>1</sub> and P<sub>1</sub>' residues Trp and Ser are conserved in all the members of the alphavirus genus.



The CP is reported to not possess trans-activity since it is an auto-proteolytic enzyme that auto-inhibits to give a turn-over number of 1 (Choi *et al.*, 1991; Hahn & Strauss, 1990). However, investigations by Morillas *et al.*, 2008 have revealed that truncations of 1-7 C-terminal residues of the SFCP, including deletion of the ultimate C-terminal tryptophan, restore the enzymatic activity of protein. The truncated enzyme was found to possess high esterase activity after deletion of the highly conserved C-terminal tryptophan. Furthermore, they concluded from the study that correct folding of the tertiary structure of SFCP is dependent on the presence of conserved Trp residue at the C-terminus (Morillas *et al.*, 2008). The findings of this study were intriguing and persuaded us to probe the structural analysis of the pre-cleavage form of another member of alphavirus genus, AVCP.

The crystal structures of the post-cleavage form of CP having the C-terminal tryptophan bound in the catalytic site have been reported from different members of the alphavirus genus including Aura virus. Additionally, the crystal structure of catalytically inactive S215A variant of Sindbis virus capsid protein (SCP) is also available, in which two additional C-terminal P<sub>1</sub>' and P<sub>2</sub>' residues (Ser265 and Ala266) are also bound near the active site (Choi *et al.*, 1996). As the structure consists of substrate residues in the specificity pocket, this structure is considered as an enzyme-substrate complex form of CP in which the catalytic triad residue Ser215 has been mutated to alanine to abolish the self-cleavage activity. Until now, the structures available for the alphavirus CP were of the inactive state with a blocked catalytic site (Choi *et al.*, 1997; Tong *et al.*, 1993; Choi *et al.*, 1991). Therefore, the development of structure-based antiviral drug design strategies targeting the CP proteolytic activity has been absent due to the unavailability of the structure of the unblocked active enzymatic form. The crystal structures of several anti-bacterial drug targets have also been proposed which further help in developing the inhibitors against these specific targets (Kumar *et al.*, 2011; Singh *et al.*, 2011; Kumar *et al.*, 2010). Some proteases are suggested to play important role as anti-bacterial and anti-obesity activity (Dave *et al.*, 2012; Mahajan *et al.*, 2012; Dave *et al.*, 2011). Furthermore, the bacterial derived cysteine protease is reported to halt the antimicrobial property of the host (Egesten *et al.*, 2009).

The C-terminal domain of CP also participates in the budding process by interacting with the glycoproteins through a hydrophobic pocket that lies on the capsid surface (Owen & Kuhn, 1997). This hydrophobic pocket has also been proposed to be involved in capsid-capsid interaction required for nucleocapsid formation (Lee *et al.*, 1996). The crystal structure of SCP containing a dioxane molecule in the hydrophobic pocket suggested the use of dioxane derivatives for targeting

## Chapter 4: Structural insights into the active form of the alphavirus capsid protease

the capsid and E2 glycoprotein interaction for antiviral development (Lee *et al.*, 1998). Some dioxane derivatives that prevent CP-glycoprotein interaction by binding to the CP hydrophobic pocket lead to defects in virus budding (Kim *et al.*, 2007; 2005). However, specific antiviral molecules that target the capsid protease activity and block the initiation of polyprotein processing have not yet been identified.

In this Chapter, the structural analysis of the C-terminally truncated protease domain from Aura virus (AVCPΔ2) is described. The high resolution atomic structure of AVCPΔ2, which represents the unblocked catalytically active form of the enzyme, has been determined. This is the first report of the crystal structure of the active form of the alphavirus CP. We discuss the conformational changes in the protein catalytic site, the oxyanion hole and the substrate specificity pockets upon transition from catalytically active to fully processed and cleaved (subsequent to auto-catalytic cleavage) inactive protein via a substrate bound complex intermediate. The observation provides insight into the properly folded structure of the trans-active form and new possibilities for structure-based antiviral drug design targeting the protease activity of alphavirus CP.

### 4.3 Methodology

#### 4.3.1 Construction of expression plasmid

The AVCPΔ2 construct (residue 110-265) was prepared in which the N-terminal disordered region and the last two residues at the C-terminus of the protease domain were deleted. Aura virus genomic cDNA was used as the template for polymerase chain reaction (PCR) amplification of the DNA fragments encoding AVCPΔ2. The oligonucleotides 5'-CTGGAATTCATATGGCCCTGAAATTTGAAGCCGAC -3' (forward) and 5'-CTAGAATCTCGAGCTATACAGTATCTTCGTGGGTGG -3' (reverse) containing *NdeI* and *XhoI* sites, respectively, were used in the PCR reaction. These primers were designed on the basis of GenBank accession no. NP\_819015.1. The PCR amplified DNA fragment was purified using the PCR purification kit (Qiagen, USA) according to the manufacturer's instructions. The purified PCR fragment and pET28c vector containing the Tobacco etch virus (TEV) protease cleavage site were digested with *NdeI* and *XhoI* restriction enzymes. The digested products were separated on a 1 % agarose gel and purified using the DNA gel extraction kit (Qiagen, USA). Restriction enzyme digested plasmid and PCR products were ligated using T4 DNA ligase. *DH5α* (DE3) cells were

transformed with the ligation mixture by the heat shock method (Inoue *et al.*, 1990). The transformed cells were plated on Luria-Bertani (LB) agar plates containing 50 µg/ml kanamycin and incubated overnight at 37 °C. The obtained colonies were picked and grown overnight at 37 °C in LB broth containing 50 µg/ml of kanamycin. Plasmids isolated from 5 ml culture using a MiniPrep plasmid isolation kit (Qiagen, USA) were screened by PCR and were assayed by restriction enzyme digestion for the presence of AVCPΔ2 insert. This insert was sequenced in both directions using T7 forward and T7 reverse primers to confirm the identity of the pET28c-AVCPΔ2 plasmid.

#### 4.3.2 Expression and Purification of AVCPΔ2

For protein production, the cloned recombinant pET28c-AVCPΔ2 plasmid was transformed into *E. coli* strain *Rossetta* (DE3). Colonies of the transformed cells were grown in LB broth supplemented with kanamycin (50 µg/ml) and chloramphenicol (35 µg/ml) at 37 °C to an optical density of 0.4 at 600 nm (OD<sub>600</sub>). At this point, the temperature was reduced to 18 °C and the culture was allowed to grow to an OD<sub>600</sub> of 0.8. Protein expression was then induced using 0.4 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and the induced culture was grown overnight at 18 °C. The cells were harvested by centrifugation and the obtained cell pellet was stored at – 80 °C till further use.

The purification procedure for AVCPΔ2 was very similar to the protocol described in Chapter 2. Briefly, the cell pellet from a 1 liter culture was re-suspended on ice in 30 ml of purification buffer (50 mM Tris-HCl pH 7.6, 15 mM imidazole and 100 mM NaCl) and the cells were disrupted using a cell disruptor (Constant Systems Ltd, Daventry, England). The cell lysate was subjected to centrifugation at 4 °C. The N-terminal His-tagged AVCPΔ2 was purified using Ni-NTA beads (Qiagen, USA) and elution was done in 250 mM imidazole. The N-terminal His-tag from AVCPΔ2 was cleaved by incubating the purified protein with TEV protease overnight and the sample was simultaneously dialyzed against the dialysis buffer (50 mM Tris-HCl pH 7.6, 20 mM NaCl) at 4 °C. After His-tag cleavage, the protein sample was reloaded onto Ni-NTA column to remove His-tagged TEV protease and uncleaved His-tagged AVCPΔ2. The flow-through containing AVCPΔ2 protein without His-tag was concentrated and loaded onto pre-equilibrated HiLoad Superdex 75 16/60 size-exclusion chromatography column (GE Healthcare) using an ÄKTA purifier (GE Healthcare), which was operated at a flow rate of 0.5 ml/min at 4 °C. Gel-filtration fractions were run on a 15 % SDS-PAGE to analyze protein purity. The fractions

## Chapter 4: Structural insights into the active form of the alphavirus capsid protease

containing pure protein were pooled and concentrated to 15 mg/ml using a 3 kDa cutoff Amicon Ultra-15 concentrator (Millipore, Bedford, Massachusetts, USA). The concentration and yield of purified protein was estimated by UV-Vis spectroscopy at 280 nm using an extinction coefficient method. The yield of AVCPΔ2 from 1 liter bacterial culture was ~ 15 mg, which is slightly higher as compared to the native AVCP.

### 4.3.3 Crystallization and data collection

Purified and concentrated protein was crystallized using the sitting drop vapor diffusion method. Protein crystallization was done in a 96 well sitting drop crystallization tray (Hampton Research, Aliso Viejo, CA) with 1 µl of protein (15 mg/ml in 50 mM Tris-HCl pH 7.6, 20 mM NaCl) and 1 µl of the reservoir solution equilibrated against 50 µl of the reservoir buffer. AVCPΔ2 crystals were obtained in 5 days using 0.2 M Sodium citrate tribasic dihydrate and 20 % w/v Polyethylene glycol (PEG) 3,350 at 20 °C. The composition of cryo-protectants was also optimized for high resolution data collection. Prior to data collection, the obtained crystal was soaked in mother liquor containing 15 % (v/v) glycerol as cryoprotectant. X-ray data were collected using a MAR 345 imaging-plate system with Cu K $\alpha$  radiation generated by a Bruker–Nonius Microstar H rotating-anode generator operated at 45 kV and 60 mA. The data were collected under cryogenic conditions at a wavelength of 1.54 Å. Diffraction data were processed and scaled using the MOSFLM package (Leslie & Powell, 2007). The data collection and processing statistics are summarized in Table 4.1.

### 4.3.4 Structure solution and refinement

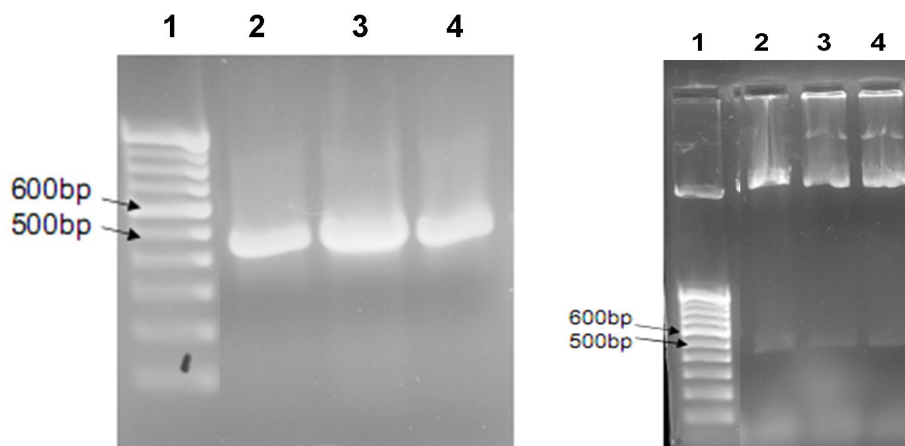
The structure determination was carried out using the molecular replacement method by taking the crystal structure of the native AVCP that has the C-terminal tryptophan residue bound to the active site as the search model. For this purpose, the MOLREP program from the CCP4 software suite was used and Refmac5 program was used for the restrained refinement (Murshudov *et al.*, 1997; Vagin & Teplyakov, 1997). The analysis of the electron density map and manual model building were carried out using the COOT program (Emsley & Cowtan, 2004). The stereochemical properties of the refined structure model of AVCPΔ2 were analyzed using the PROCHECK program (Laskowski *et al.*, 1993). Structural analysis of the refined model and the preparation of figures were done using the PyMOL visualization tool (Delano, 2008). For the

analysis of dimer interface contacts, the PISA (Protein Interfaces, Surfaces and Assemblies) web server was used (Krissinel & Henrick, 2007).

## 4.4 Results and Discussion

### 4.4.1 Purification of active AVCP $\Delta$ 2

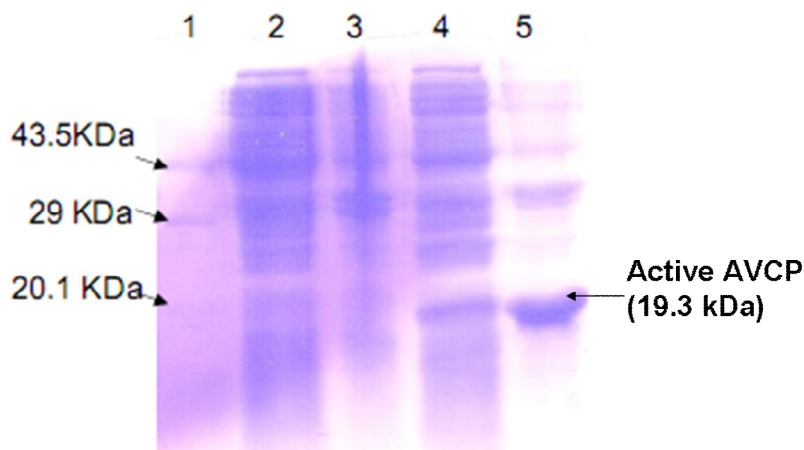
The carboxyl terminal Trp267 of AVCP remains bound to the S<sub>1</sub> specificity pocket blocking the entry of substrate for trans-cleavage. This blocked enzyme represents the post-cleavage form of CP. Deletion of the conserved Trp267 is anticipated to free the blocked active site, making it accessible to the peptide substrate for trans-protease cleavage. This trans-active protease would represent the pre-cleavage state of auto-proteolytic alphavirus CP. Therefore, the AVCP $\Delta$ 2 (110-265) construct with deleted C-terminal Glu266 and Trp267 residues from the AVCP domain (110-267) was designed and cloned in a bacterial expression vector pET28c having the TEV protease cleavage site after the histidine tag. The restriction sites used in the cloning of the active AVCP construct were chosen accordingly so that the His tag coding gene remains in the vector along with AVCP encoding gene. The presence of the His tag at the N-terminus of the protein would further facilitate protein purification. The pET28c-AVCP $\Delta$ 2 plasmid contains the His tag, TEV protease cleavage site and the gene encoding the AVCP with deleted disordered N-terminal region and last C-terminus residues. The agarose gel showing the PCR product and the confirmation of the cloning by restriction enzyme digestion is shown in Figure 4.4.1.1.



**Figure 4.4.1.1:** Cloning of AVCP $\Delta$ 2 construct (a) PCR amplification, Lane 1 DNA ladder, Lane 2, 3 & 4 PCR (110-265); (b) Restriction digestion confirmation of the cloned gene, Lane 1 DNA ladder, Lane 2, 3 & 4 RE digestion (110-265).

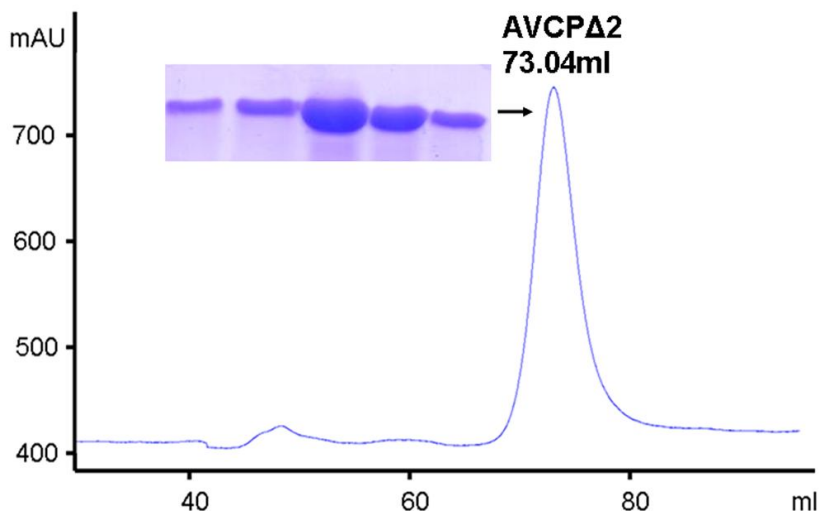
#### Chapter 4: Structural insights into the active form of the alphavirus capsid protease

The recombinant AVCP $\Delta$ 2 protein having a TEV protease removable 6x His affinity tag at the N-terminus was produced in soluble form in *E. coli*. The protein expression was optimized for IPTG concentration, induction time and induction temperature. The protein expressed in the soluble form is shown in Figure 4.4.1.2



**Figure 4.4.1.2:** Expression of AVCP $\Delta$ 2 construct, Lane 1 Protein mol. wt. Marker, Lane 2 Control Supernatant, Lane 3 Control Pellet, Lane 4 Induced Supernatant, Lane 5 Induced Pellet.

Protein from the soluble fraction was purified using Ni<sup>2+</sup> affinity and size exclusion chromatography. The N-terminal His<sub>6</sub>-tag was removed from the purified protein using TEV protease. SDS-PAGE analysis showed the presence of a single protein band of ~ 17 kDa confirming sample purity and homogeneity of purified AVCP $\Delta$ 2 (Figure 4.4.1.3).

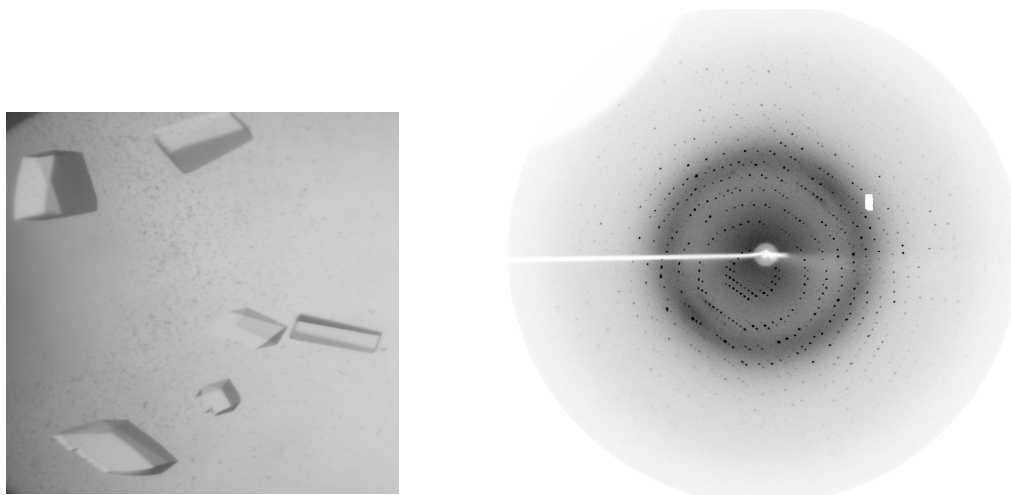


**Figure 4.4.1.3:** Gel filtration chromatography and SDS-PAGE analysis of major peak fractions show the protein (AVCP $\Delta$ 2) purified to homogeneity. The protein is monomeric in nature as determined by gel filtration profile.

A standard curve using gel-filtration molecular weight markers was prepared and the average molecular mass of the major elution peak of purified protein was calculated using the calibration curve. The estimated molecular weight of the protein was calculated to be ~ 17 kDa suggesting that AVCPΔ2 is a monomer in solution.

#### 4.4.2 Crystallization of AVCPΔ2

To determine the tertiary structure of AVCPΔ2, the purified protein was concentrated to ~ 15 mg/ml and crystallization attempts were made. After a number of trials, diffraction quality crystals were obtained by using 0.2 M Sodium citrate tribasic dihydrate and 20 % w/v Polyethylene glycol 3,350 at 20 °C (Figure 4.4.2.1). From single crystal, a complete data set was collected at 1.81 Å (Table 4.1). The data were collected at home source and processed further. The diffraction pattern is shown in Figure 4.4.2.1. AVCPΔ2 crystals belong to monoclinic lattice with space group  $P2_1$ , however, inactive AVCP crystals reported earlier also belonged to the monoclinic lattice but in space group  $C2$ . Two molecules per asymmetric unit are found in AVCPΔ2 whereas inactive AVCP has one monomer per asymmetric unit. The estimated Matthews coefficient was  $3.15 \text{ \AA}^3 \text{ Da}^{-1}$  and solvent content was 60.97 % for two molecules of AVCPΔ2 per asymmetric unit.



**Figure 4.4.2.1:** Crystals obtained for the active form of AVCP and the diffraction pattern.

**Table 4.1:** Data collection and refinement statistics for active form of AVCP.

	Active AVCP
Crystallographic Data	
Space group	<i>P21</i>
Cell dimensions (Å)	<i>a</i> =53.56, <i>b</i> =71.35, <i>c</i> =61.51 <i>α</i> =90.00, <i>β</i> =114.33, <i>γ</i> =90.00
Resolution range (Å)	56.08 – 1.81
Completeness (%) (Last Shell)	97.6 (83.6)
<i>R</i> <sub>merge</sub> † (%) (Last Shell)	5.6 (51.0)
Mean <i>I</i> / <i>σ</i> ( <i>I</i> ) (Last Shell)	15.1 (3.0)
No. of observed reflections (Last Shell)	147227 (17067)
No. of unique reflections (Last Shell)	38673 (4783)
Molecules per asymmetric unit	2
Matthews coefficient (Å <sup>3</sup> Da <sup>-1</sup> )	3.15
Solvent content (%)	60.97
Multiplicity (Last Shell)	3.8 (3.6)
Refinement	
No. of Residues	306
Water molecules	400
<i>R</i> <sub>cryst</sub> (%)	18.5
<i>R</i> <sub>free</sub> (%)	21.8
Average <i>B</i> -factor (Å <sup>2</sup> )	23.01
r.m.s.d on bond lengths (Å)	0.013
r.m.s.d on bond angles (Å)	1.434
Ramachandran plot (%)	
Most Favoured	95.5
Allowed	4.5
Outliers	0.0

†  $R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - [I(hkl)]|}{\sum_{hkl} \sum_i I_i(hkl)}$ , where  $I_i(hkl)$  is the *i*th observation of reflection *hkl* and  $[I(hkl)]$  is the weighted average intensity for all observations *i* of reflection *hkl*.

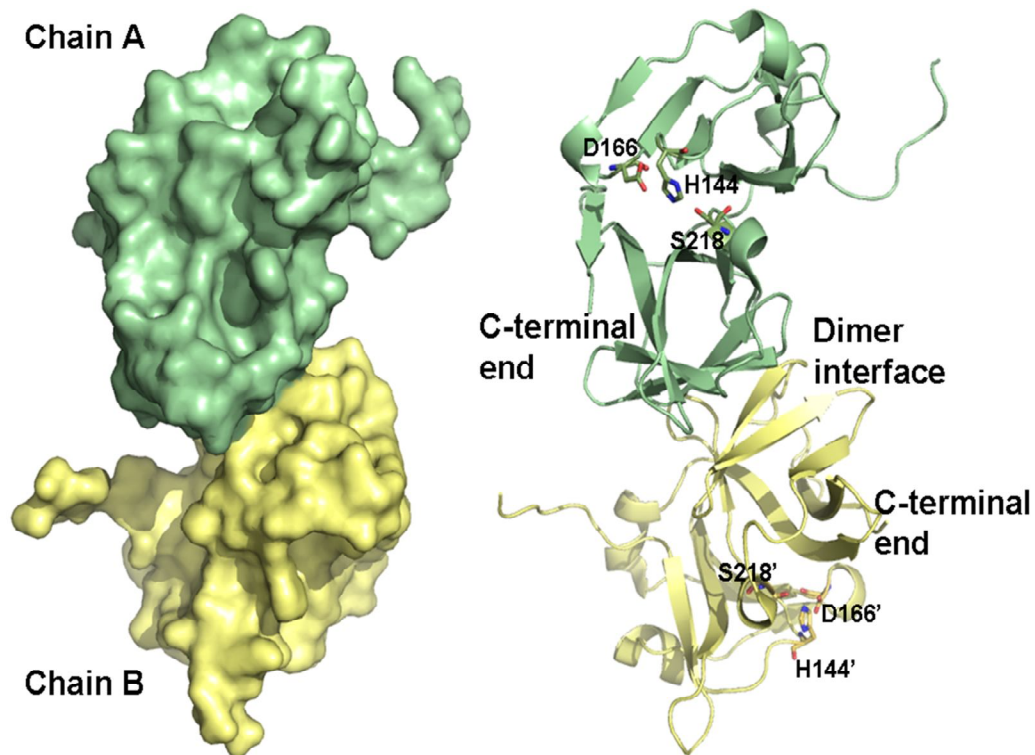


#### 4.4.3 Overall structure of AVCP $\Delta$ 2

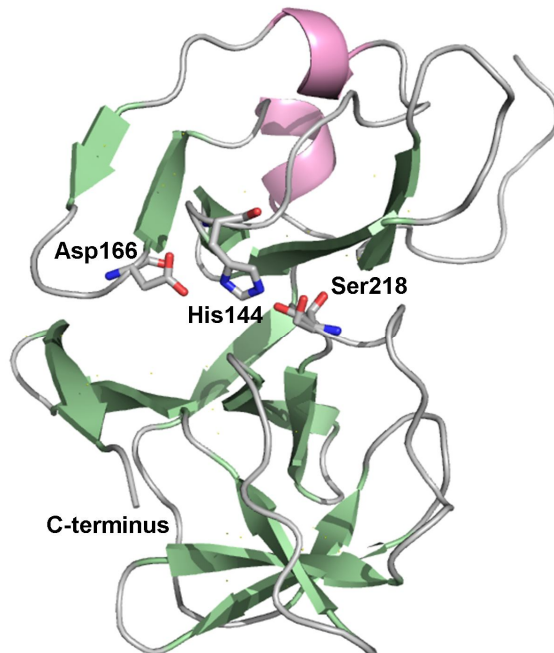
The structure of AVCP $\Delta$ 2 was determined at 1.81 Å resolution and refined to a final Rfactor of 18.5 % and Rfree of 21.8 % with more than 95 % residues in the most favored region of the Ramachandran plot. High resolution data show clear electron density for the N-terminal six residues, which were disordered and thus not visible in the native AVCP crystal structure.

Similar to the other already known chymotrypsin-like serine proteases, the structure of each monomer of AVCP $\Delta$ 2 consists of two sub-domains with each domain made up of 6-7 stranded, anti-parallel,  $\beta$ -barrel structure. The surface and the cartoon view of the overall structure of the active AVCP is shown in Figure 4.4.3.1. The structure shows the presence of crystallographic dimer in active AVCP, however, the protein is monomeric in solution. The catalytic triad consisting of residues His144, Asp166 and Ser218 is placed at the interface of these two domains. The monomer of AVCP $\Delta$ 2 consists of thirteen antiparallel  $\beta$ -sheets [six in domain I formed by residues 118-122 ( $\beta$ 1), 128-135 ( $\beta$ 2), 138-142 ( $\beta$ 3), 148-149 ( $\beta$ 4), 160-162 ( $\beta$ 5) and 167-171 ( $\beta$ 6); and seven in domain II by residues 190-194 ( $\beta$ 7), 197-202 ( $\beta$ 8), 205-209 ( $\beta$ 9), 221-223 ( $\beta$ 10), 229-238 ( $\beta$ 11), 242-250 ( $\beta$ 12) and 256-259 ( $\beta$ 13)] and two  $3_{10}$  helices [formed by residues 153-156 ( $\alpha$ 1) and 174-177 ( $\alpha$ 2)]. Both of these  $3_{10}$  helices are present in domain I.

The overall fold within each monomer of AVCP $\Delta$ 2 is very similar to the native AVCP with root mean square deviation (RMSD) of 0.33 Å on C $\alpha$  atoms. However, the RMSD on superimposition of the native AVCP and native AVCP with bound dioxane in the E2 binding hydrophobic pocket of CP was only 0.13 Å. This indicates that the main-chain conformational variations are more pronounced between the native and active form as compared to the two forms (apo and ligand bound) of native AVCP. The crystal structure of a single monomer of active AVCP is represented in Figure 4.4.3.2.

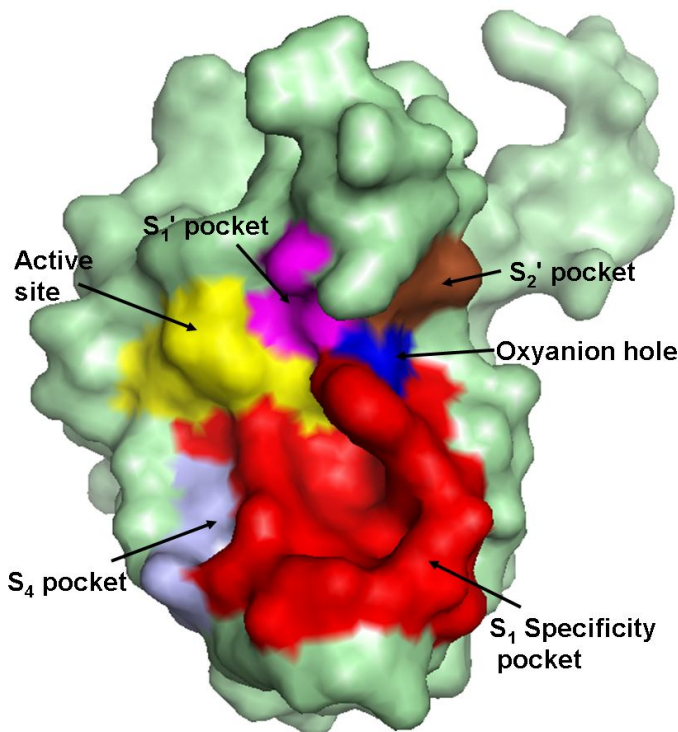


**Figure 4.4.3.1:** Overall structure of active AVCPΔ2 construct [surface view (left) and the cartoon form (right)] showing the dimeric form and the catalytic triad residues. Chain A is shown in green while chain B is in yellow. The catalytic triad residues are represented as sticks.



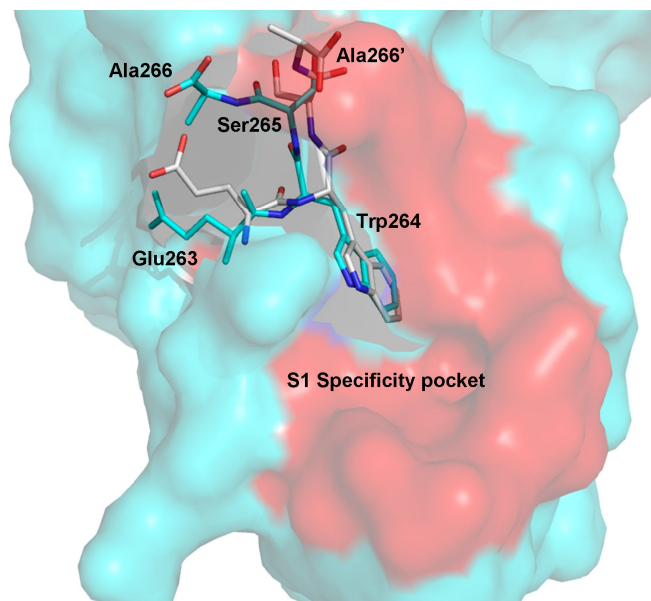
**Figure 4.4.3.2:** Overall structure of the monomer contains two  $\beta$ -barrel subdomains consisting of the catalytic triad in between the cleft. The catalytic triad residues are presented in sticks. The coloring is done on the basis of secondary structure.

The surface view of the monomer depicting different specificity pockets, the oxyanion hole and the active site is shown in Figure 4.4.3.3. The specificity pockets are responsible for the substrate specificity of the enzyme. The structural complementarities of both the substrate and the substrate binding pocket allow proper binding of the substrate to get cleaved by the enzyme.



**Figure 4.4.3.3:**Surface view of AVCPΔ2 showing different pockets and regions involved in catalysis, coded with different colors.

At the C-terminus of AVCPΔ2, the electron density for the last four residues (262-265) is missing which indicates conformational flexibility of these residues in the absence of the last two residues including Trp. The two chains of the substrate bound mutant SCP structure shows very high flexibility of this region. In Chain A, the substrate is not able to reach the active site and thus considered as early enzymatic state with loosely bound substrate. However, the C-terminus of Chain B enters the S<sub>1</sub> specificity pocket and is found near the active site similar to the chymotrypsin structure with substrate bound to it (Choi *et al.*, 1996). Figure 4.4.3.4 clearly depicts the differences in the substrate binding for the two Chains in SCP mutant form.

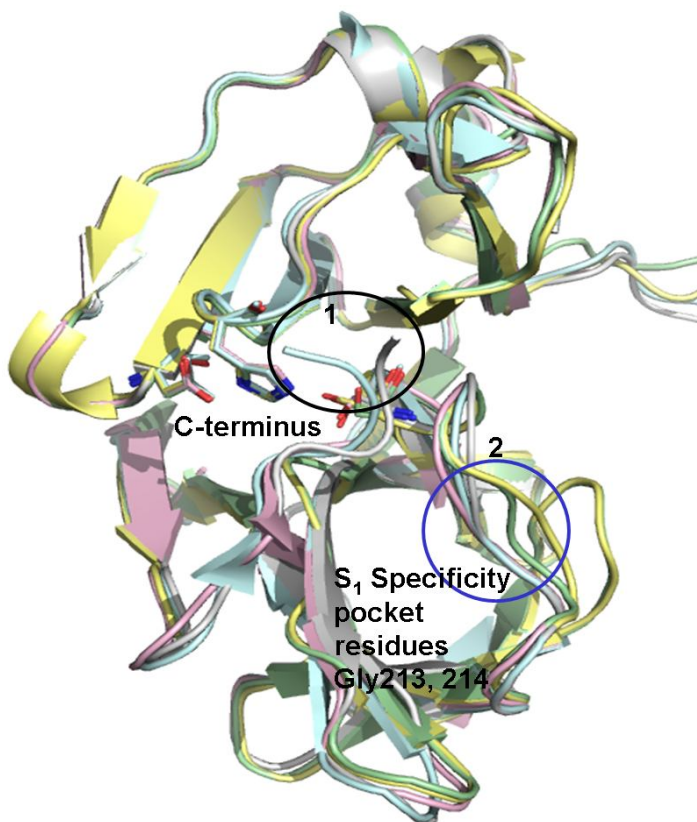


**Figure 4.4.3.4:** The  $S_1$  specificity pocket of SCP substrate bound mutant form (PDB ID: 1SVP). The specificity pocket is shown as surface view with the specificity pocket residues in red. The superimposition of the substrate residues from both the chains of SCP substrate bound mutant form is done. The chain A is shown in blue color while the chain B is in gray. The substrate residues are shown as sticks. A large conformational variation is present between the substrate residues of the two chains. In chain B, the substrate shows proper binding to the specificity pocket while in chain A, it is present away from the specificity pocket as well as the active site. Thus, Chain A has been considered as early enzymatic state with loosely bound substrate.

In native AVCP, the C-terminal loop takes a  $\beta$  turn at residue His261 to allow Trp267 to bind the substrate specificity pocket. In AVCP $\Delta 2$ , Trp267 is truncated so the loop movement is flexible and no longer restricted due to the binding of Trp267 at the  $S_1$  specificity pocket. The movement and flexibility of loop is crucial for proper positioning of the scissile peptide bond relative to the active site and oxyanion hole for cleavage. Thus, this loop movement towards the active site seems to play an important role in the binding of Trp267 to the active site and the auto-proteolytic property of CP. The C-terminal loop, which is missing from unblocked form and found in the Trp bound native state, shows different conformations in two chains of the SCP substrate bound intermediate form (Figure 4.4.3.4). This suggests that the loop is highly flexible in three different enzymatic states and becomes restrained in position after proteolytic cleavage.

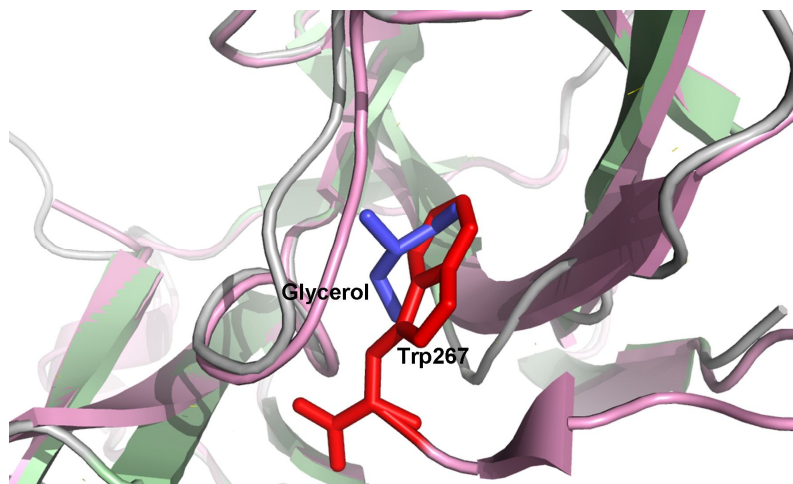
The structural alignment of the native AVCP, both chains of active AVCP $\Delta 2$  and SCP substrate bound mutant shows clear structural differences in different enzymatic forms. The loop regions show the flexibility in between different forms. The loop region present at the  $S_1$  specificity pocket is highly flexible among different stages of the CP. However, the catalytic triad residues remain almost at the same position in all the different forms. At some places there is a

difference in the secondary structure of the CP. The structures of the CP from other members of alphavirus genus also have the similar overall structure. The alignment of different forms of the CP structure is presented in Figure 4.4.3.5 which shows the remarkable difference at the N and C terminus of the capsid structure in different forms.



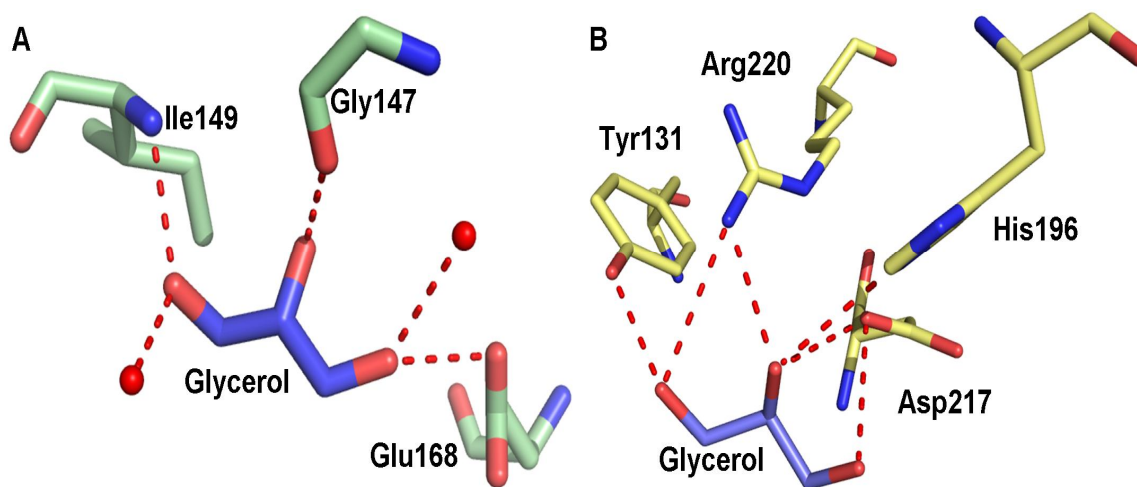
**Figure 4.4.3.5.** Both chains of active (Chain A in green and Chain B in yellow color) and substrate-bound form (Chain A in blue and Chain B in gray) were aligned along with native AVCP (pink). Circle 1 (black) shows the variation in the C-terminal region in two chains of substrate bound form, while this region is absent from the active AVCP and remains intact in native AVCP. Circle 2 (blue) shows the loop flexibility in S<sub>1</sub> specificity pocket.

In AVCP $\Delta$ 2, the structure of active form of the enzyme, three glycerol molecules (used as cryoprotectant) are clearly visible in the electron density map of both the chains and show interactions with different amino acid residues. Superposition of the crystal structure of AVCP in the inactive state, and AVCP $\Delta$ 2 the active form, shows that a glycerol molecule binds exactly at the same position where Trp267 is present in the S<sub>1</sub> specificity pocket of AVCP (Figure 4.4.3.6).



**Figure 4.4.3.6:** The superposition of the active and native AVCP shows the presence of one glycerol molecule (blue in color) exactly at the same position where Trp267 (red in color) side chain lies in the specificity pocket.

Another glycerol molecule in both chains of AVCP $\Delta$ 2 occupies a cavity by making polar interactions with the main chain of Gly147 and Ile149, side chain of Glu168 and two water molecules. Gly214 in chain A and Tyr131, His196, Asp217 and Arg220 in chain B interact with the third glycerol molecule through H-bonding (Figure 4.4.3.7). As glycerol has been proven to be a protein structure stabilizing agent, hence glycerol used in soaking the crystal before cryo-freezing and data collection might be influencing the structure of AVCP $\Delta$ 2 through stabilization (Farnum & Zukoski, 1999; Sousa, 1995; Gekko & Timasheff, 1981). Few other cases are reported in which the organic solvent help in stabilizing the protein structure (Dave *et al.*, 2011).



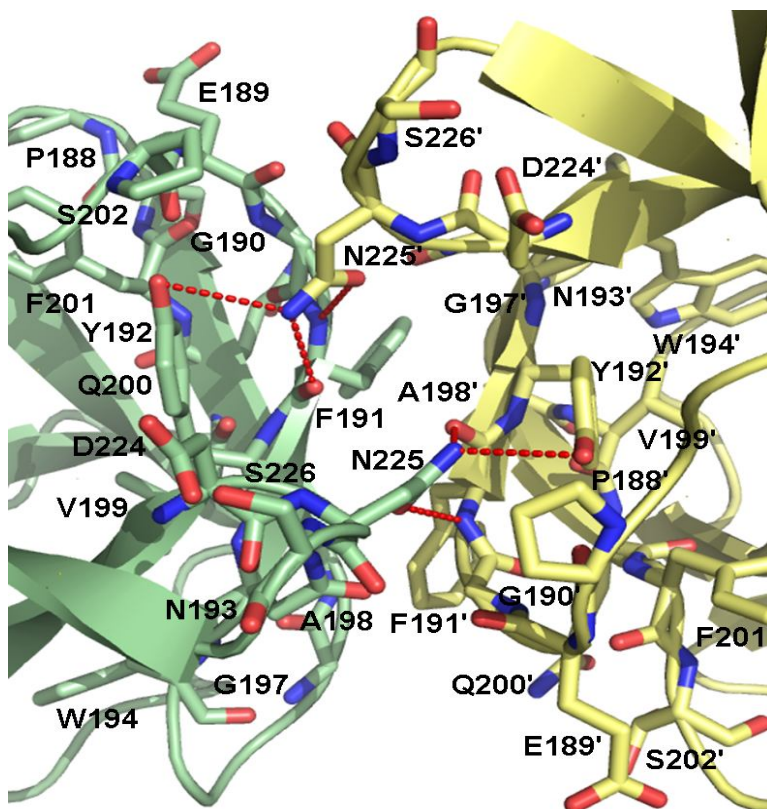
**Figure 4.4.3.7:** The interaction of the glycerol with different residues is shown. The glycerol molecule and the residues are shown in sticks. The water molecules that interact with the glycerol are presented as red spheres. (A) shows the glycerol molecule bound to the Chain A (green sticks) while (B) depicts Chain B residues (yellow sticks) with bound glycerol.

#### 4.4.4 Dimeric interactions

This study shows the monomeric nature of protein in solution and the crystallographic dimer in the asymmetric unit. In the AVCP $\Delta$ 2 structure, both the monomers are in a tail-to-tail contact through their C-terminal sub-domains. Previously, a number of dimers and trimers were found in the SCP and the SFV CP that were arranged in head-to-tail and tail-to-tail fashion in different crystal forms (Choi *et al.*, 1997; Choi *et al.*, 1996; Choi *et al.*, 1991). In head-to-tail dimers, the N- and C-terminal sub-domains are in contact as in the SCP S215A mutant crystal structure (Choi *et al.*, 1996). In head-to-head dimer formation, the N-terminal sub-domains are in contact as in the SFV CP wild-type I crystal type (Choi *et al.*, 1997).

The dimeric interface area in AVCP $\Delta$ 2 as calculated using the PISA web server was found to be  $\sim 375 \text{ \AA}^2$ , comprising approximately 4.5 % of the total solvent accessible area for each monomer that was consistent with previous reports in SCP (Choi *et al.*, 1996; Tong *et al.*, 1993). The dimer interaction involves two sheets VII and VIII (190GFYNW194 and 197GAVQFS202) and two loops (188PE189 and 224DNS226) of both the monomers. The two chains are held together by a few hydrogen bonds and mostly non-bonded interactions (Choi *et al.*, 1997; Choi *et al.*, 1996).

These contacts are facilitated by the contribution of 11 residues from each subunit that are common to both the monomers. The residues Pro188, Glu189, Gly190, Phe191, Tyr192, Asn193, Gly197, Ala198, Asp224, Asn225 and Ser226 from both monomers comprise the dimer interface. These mainly involve hydrophobic interactions along with hydrogen bond formation by residue Asn225 of one monomer with Phe191 and Tyr192 residues of the other monomer (Figure 4.4.4.1). The hydrophobic interactions were found to play a more important role in dimer formation as compared to the hydrogen bonding. However, no evidence is available for the biological relevance of this dimer. Moreover, the mutations that lead to crystallographic dimer disruption show no adverse effect on viral replication (Choi *et al.*, 1996). Thus, two molecules in the asymmetric unit do not represent a physiological dimer and this is consistent with the AVCP being a monomer in solution.

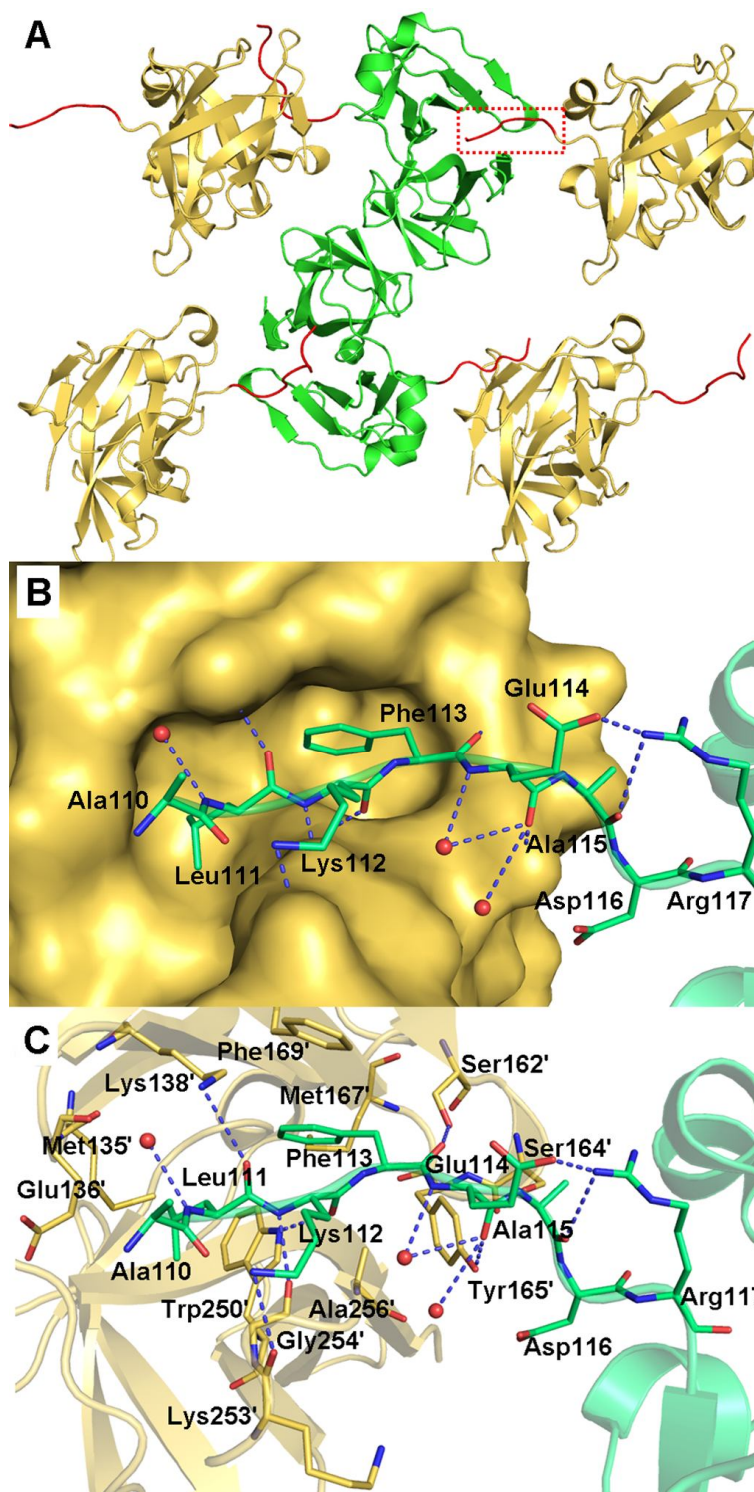


**Figure 4.4.4.1:** Close-up view of the dimeric interaction in AVCP $\Delta$ 2. The residues involved in crystallographic dimer formation are shown in sticks. Asn225 of one chain forms H-bonds with Phe191 and Tyr192 of the other chain. Chain B residues are labeled with an apostrophe sign. Chain A is shown in green while chain B is in yellow color.

#### 4.4.5 Conformational switching in the hydrophobic pocket

The electron density corresponding to the N-terminal residues from 110-115 (plus two extra residues at the N-terminal left after His-tag cleavage) is present in both chains of the AVCP $\Delta$ 2 structure. These residues were disordered in the previously reported inactive AVCP and in the dioxane-bound AVCP structures. These residues in AVCP $\Delta$ 2 form an arm that projects out of the structure. A detailed analysis of AVCP $\Delta$ 2 structure shows that this N-terminal arm interacts and binds in the hydrophobic pocket which is formed between the two  $\beta$ -barrel sub-domains of the neighboring symmetry monomer molecule (Figure 4.4.5.1a & b). The hydrophobic residues Leu111 and Phe113 of the N-terminal arm bind at the pocket between Tyr183, Trp250 and Phe169. Other residues that line this cavity are Met135, Glu136, Lys138, Ser162, and Met167. Figure 4.4.5.1c shows various interactions that stabilize the binding of the N-terminal arm in this pocket. Earlier crystallographic reports of the alphavirus capsid have also reported that this hydrophobic pocket is occupied by similar binding of the hydrophobic N-terminal arm residues from neighboring CP molecules (Choi *et al.*, 1996; Lee *et al.*, 1996).





**Figure 4.4.5.1: The N-terminus of CP bound to the hydrophobic pocket of neighboring CP molecule.**  
**a:** The crystal structure of AVCP $\Delta$ 2 dimer (green in color) with the symmetry related molecules (brown in color). The hydrophobic pocket of each molecule occupied with the N-terminal arm (red in color) of the neighboring symmetry related molecule is shown. The red color dotted square highlights one such binding.  
**b:** The zoomed view of hydrophobic pocket of one of the AVCP $\Delta$ 2 subunit with bound N-terminal arm (green in color) of the neighboring subunit is displayed in the surface view and (c) cartoon view. All the polar interactions are shown with blue dotted lines and the interacting water molecules are shown as red color spheres.

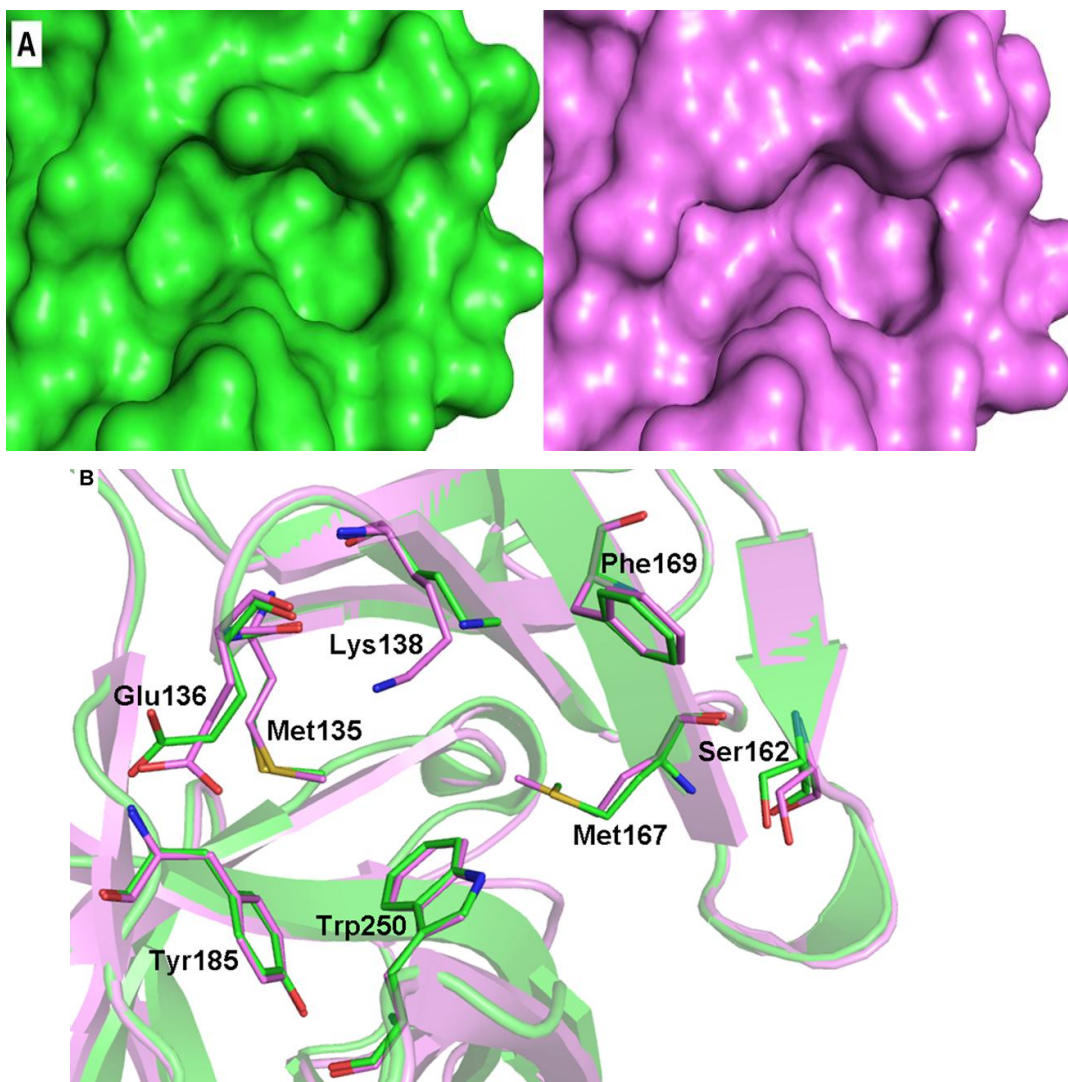
#### Chapter 4: Structural insights into the active form of the alphavirus capsid protease

Based on structural and mutational studies, it is hypothesized that this hydrophobic pocket plays a role at two stages in the virus life cycle. Firstly, the binding of the N-terminal arm residues to the hydrophobic pocket in crystal structures of two different alphaviruses including AVCP $\Delta$ 2 and SCP suggests a biological significance of this interaction (Choi *et al.*, 1996; Lee *et al.*, 1996). Furthermore, mutational studies of capsid residues 108 and 110 in the SCP have revealed the role of this interaction in capsid assembly (Lee *et al.*, 1996). Thus, it is proposed that binding of the N-terminal arm in the hydrophobic pocket of neighboring CP contributes to the formation of nucleocapsid core by linking capsid proteins together before RNA binding and core assembly.

Additionally, sequence analogy of cytoplasmic domain of E2 (cdE2) and the N-terminal arm of CP, mutational, modeling and structural studies have suggested that cdE2 and CP hydrophobic pocket interactions are crucial for virus budding process (Lee *et al.*, 1998; Owen & Kuhn, 1997). A conformational switching in the hydrophobic pocket has been proposed that changes the role of this pocket from core assembly to virus budding (Choi *et al.*, 1997).

The structural comparison of the hydrophobic pocket in the pre-cleavage state AVCP $\Delta$ 2 (with bound N-terminal arm) and the post-cleavage state AVCP (with unoccupied hydrophobic pocket) visibly illustrate conformational changes in the hydrophobic pocket (Figure 4.4.5.2a). It is presumed that the slightly wider pocket with the bound N-terminal arm in AVCP $\Delta$ 2 represents the conformation likely to promote core assembly. Whereas the narrower pocket with somewhat two distinct sub-pockets in inactive AVCP is perhaps engaged in interaction with cdE2 desirable for virus budding. As shown in Figure 4.4.5.2b, minor variations in most of the side chains of residues (Glu136, Met135, Lys138, Ser162, Met167, and Phe169) lining the hydrophobic pocket direct this conformational variation. However, further investigations are required to confirm this presumption of conformational switching in the hydrophobic pocket.

The conformational rearrangements are observed by comparison of the pre-cleavage and the post-cleavage states of AVCP in the CP hydrophobic pocket (Owen & Kuhn, 1997). The conformational changes may be responsible for the switching mechanism of the hydrophobic pocket in binding to N-terminal arm for core assembly and then its interaction with cdE2 glycoprotein during virus budding.



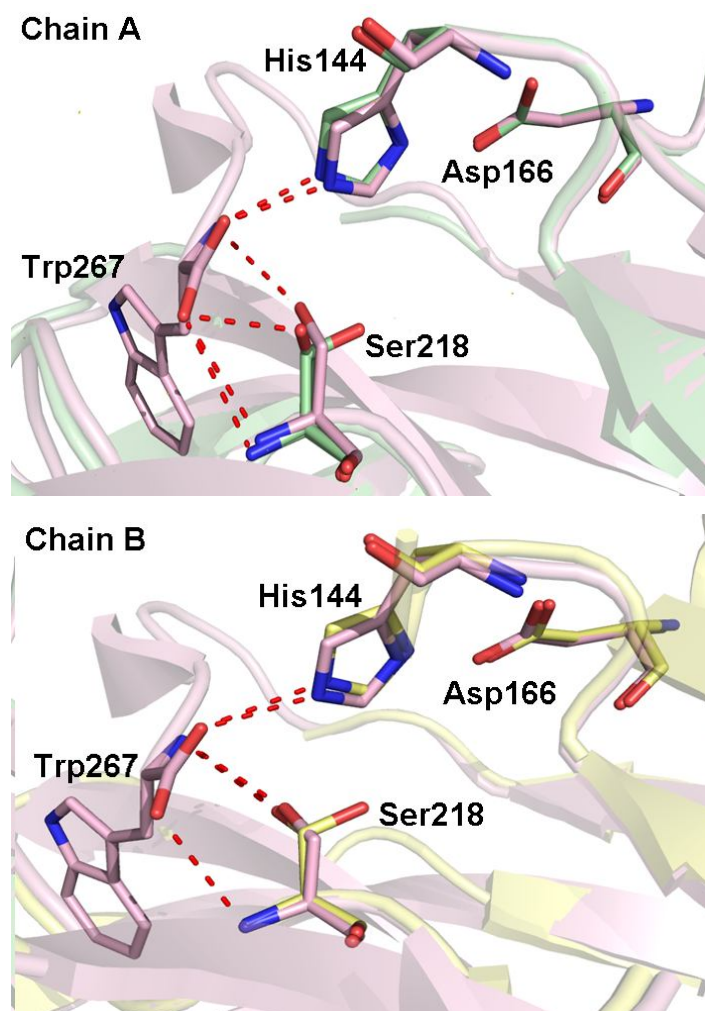
**Figure 4.4.5.2a:** The surface view and **b:** the superposed view of the hydrophobic pocket of AVCPΔ2 (green in color) over AVCP (pink in color) crystal structure to show the differences in the side chain conformation of some of the residues that occur at this hydrophobic pocket to accommodate the N-terminal arm of the neighboring subunit.

#### 4.4.6 The catalytic triad

The alphavirus CP has a fold similar to the chymotrypsin-like serine proteases and the conserved catalytic triad residues Ser218, His144 and Asp166 of AVCP form the active site at the interface between the two  $\beta$ -barrel sub-domains (Figure 4.4.3.1) (Choi *et al.*, 1997; Choi *et al.*, 1996; Tong *et al.*, 1993; Choi *et al.*, 1991). The polar interactions of Ser218 and His144 with Trp267, and the complementary shape of the  $S_1$  specificity pocket and side chain of Trp267 likely fix the position of the completely conserved  $P_1$  residue Trp267 in the active site for scissile bond cleavage. In both chains of the AVCPΔ2 structure, which represents the pre-cleavage state, Ser218 is found to be present in two alternate conformations showing its flexibility, and gets fixed upon

#### Chapter 4: Structural insights into the active form of the alphavirus capsid protease

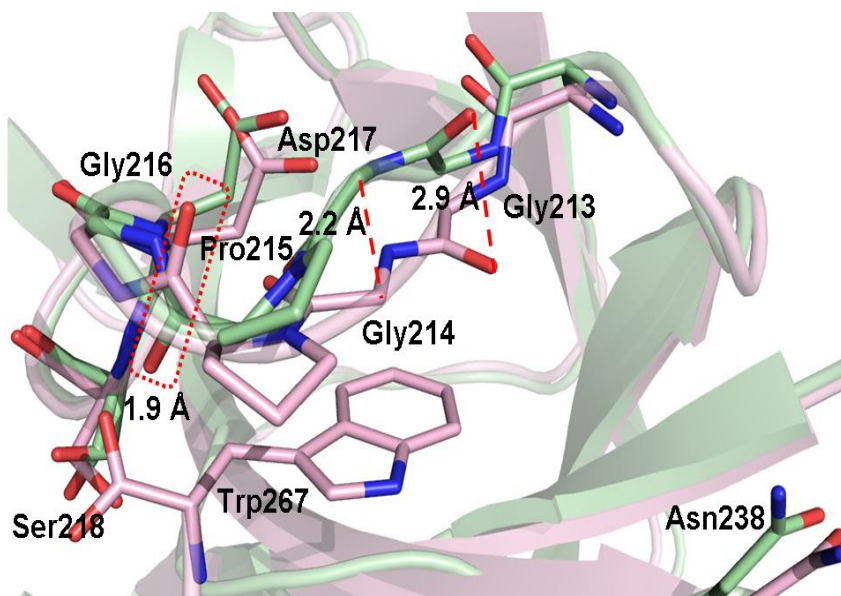
entry of substrate in native AVCP (Figure 4.4.6.1). However in the post-cleavage state, represented by AVCP, Ser218 does not show flexibility as the carboxyl terminal Trp267 produced on scissile bond cleavage makes polar contacts with the Ser218 side chain. The Ser218 main chain shows displacement of only 0.4 Å between the active pre-cleavage and the inactive post-cleavage states. His144 and Asp166 show similar polar interactions with each other and with no alternate conformations in various enzymatic states. Only a minor difference of 0.4-0.6 Å is detectable in the side chain of His144. The Asp166 residue is partially exposed to the solvent and interacts with two conserved water molecules in both enzymatic forms.



**Figure 4.4.6.1:** The catalytic triad residues from different enzymatic states: native AVCP (pink), AVCP $\Delta$ 2 Chain A (green) and AVCP $\Delta$ 2 Chain B (yellow) show conformational change in Ser218 side chain. The catalytic triad residues and Trp267 are shown as sticks. The interactions of Ser218 and His144 with Trp267 are shown as red dash line.

#### 4.4.7 The oxyanion hole

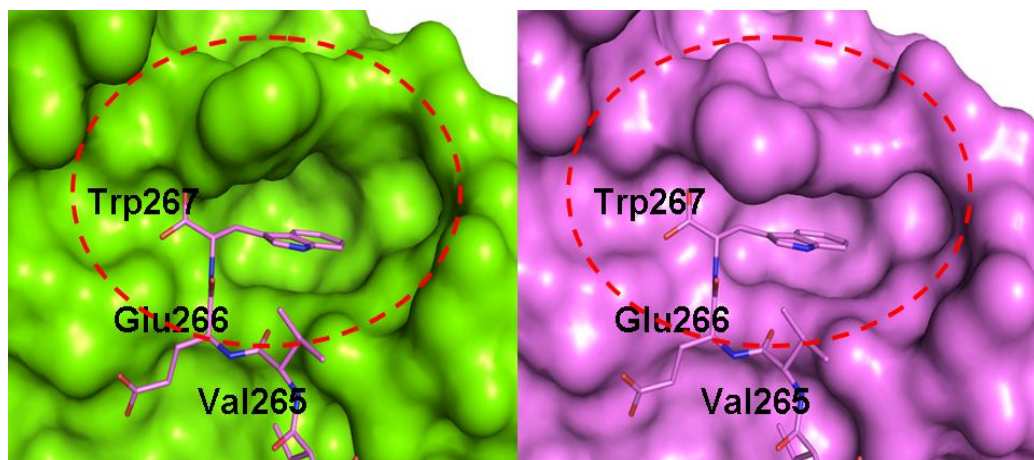
The oxyanion hole, indispensable for proteolytic activity, plays a role in the substrate binding and stabilization of the tetrahedral transition intermediate containing the oxyanion through polar interactions with the main chain amides of the conserved GDSG motif (Rumthao *et al.*, 2004; Cui *et al.*, 2002). Interestingly, the crystal structure of AVCP $\Delta$ 2 shows altered conformation for its oxyanion hole (Figure 4.4.3.4). This has resulted due to a flip of the peptide bond between Pro215 and Gly216 as shown by Figure 4.4.7.1. The oxyanion hole residues Gly216 and Gly219 show backbone conformational differences in both structure forms. A displacement of up to 2.9 Å can be observed in Gly213-Gly216 as shown in Figure 4.4.7.1. Thus, it reveals the role of the oxyanion hole main chain in substrate binding. The Ser218 residue conformational change does not impart any effect on the alteration of the oxyanion hole conformation as can be seen from the mutant S215A crystal structure of SCP and an equivalent substitution in subtilisin (Choi *et al.*, 1996; Carter & Wells, 1988). Superimposition of crystal structures of AVCP $\Delta$ 2, the active pre-cleavage state and AVCP, the inactive post-cleavage state, shows that the carboxylate oxygen atom of Trp267 in AVCP is just 1.9 Å away from the carboxyl oxygen of Pro215 in AVCP $\Delta$ 2 indicative of steric hindrance in substrate binding.



**Figure 4.4.7.1:** The differences at the  $S_1$  specificity pocket of AVCP $\Delta$ 2 (green in color) are shown in superposition with native AVCP (pink in color) crystal structure with bound Trp267 at the active site. The dotted yellow lines highlight the difference at region Gly213-Gly216. The red colored dotted box highlights the flipped peptide bond at Pro215 and Gly216.

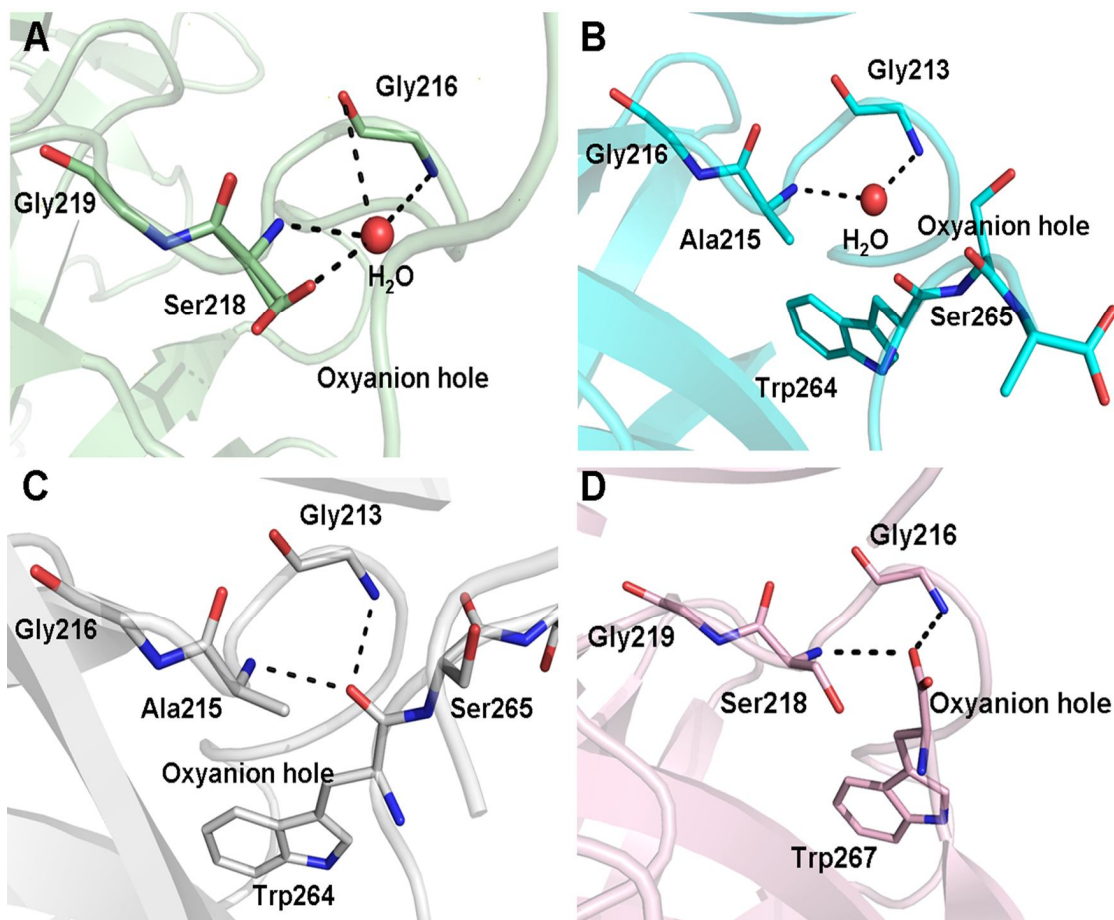
These conformational changes in residues from Gly213-Gly217 change the overall geometry of the specificity pocket as shown in the surface view in figure 4.4.7.2. Hence, it can be

hypothesized that region 213-217 has an altered conformation in the absence of substrate in the specificity pocket of the pre-cleaved state and changes during the substrate binding to prevent the steric hindrance between Pro215 and the incoming substrate.



**Figure 4.4.7.2:** The differences in the volume of  $S_1$  specificity pocket are shown in surface view.

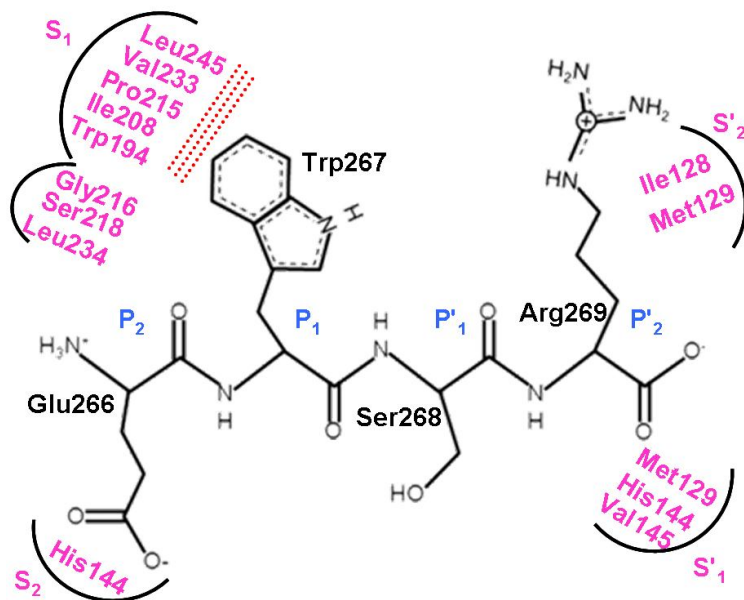
The active form, AVCP $\Delta$ 2 contains a water molecule in the oxyanion hole (Figure 4.4.7.3a). Interestingly, this water molecule is conserved in Chain A of the SCP substrate bound mutant S215A, which is considered as an intermediate form (Figure 4.4.7.3b), whereas in its Chain B, Gly213 (corresponds to Gly216 in AVCP) main chain NH atom forms an H-bond with carbonyl oxygen atom preceding scissile bond (Figure 4.4.7.3c). This interaction was absent in Chain A of SCP substrate bound mutant S215A as the loop that brings the substrate inside the substrate binding pocket is highly flexible and doesn't reach the oxyanion hole in Chain A. Post-cleavage AVCP also doesn't contain water molecules in the oxyanion hole and shows similar interactions with the C-terminal carboxylate as that of Chain B of the SCP substrate-bound form (Figure 4.4.7.3d). Previously, Chain A of substrate bound form in SCP was considered as early enzymatic state in which substrate is loosely bound and is more similar to chymotrypsin without substrate (Choi *et al.*, 1996). Hence, the unblocked active AVCP $\Delta$ 2 represents the pre-cleavage state of the alphavirus CP prior to cis-catalytic activity. Consequently, it can be concluded that the substrate scissile bond replaces a water molecule present in the oxyanion hole of AVCP $\Delta$ 2 during proteolysis.



**Figure 4.4.7.3:** (a & b) Oxyanion hole of AVCP $\Delta$ 2 (green) and Chain A of substrate bound SCP (blue) occupy a water molecule. However, (c & d) represent the H-bonding of oxyanion hole residues with carbonyl oxygen atom preceding scissile bond in chain B of substrate bound form and native AVCP respectively.

#### 4.4.8 The substrate specificity pockets

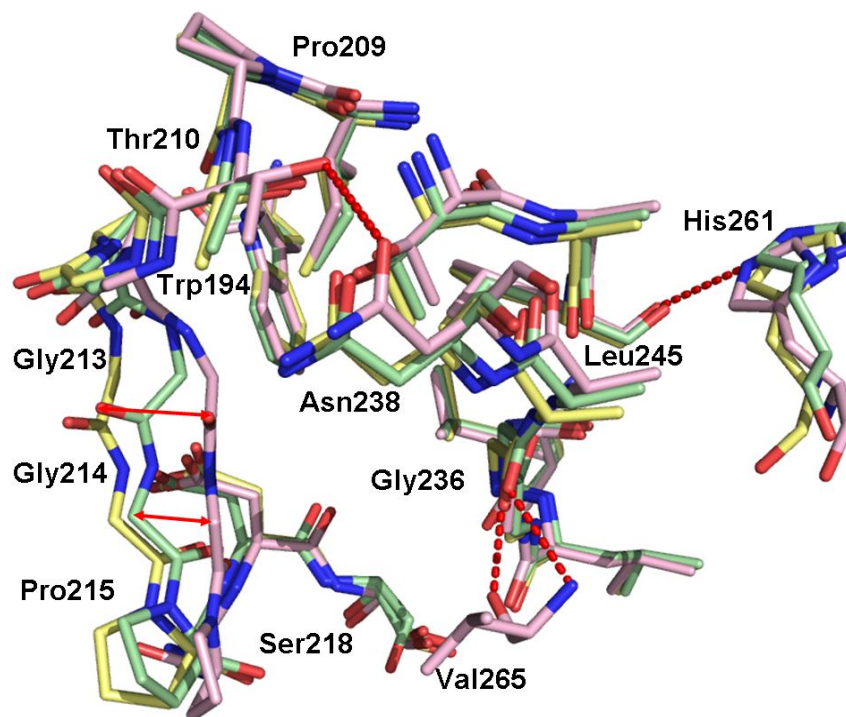
For comparative structural analysis, the structure of AVCP $\Delta$ 2 was superimposed onto the inactive AVCP as well as onto the substrate bound SCP mutant S215A structures. A series of conformational changes to variable extent in different specificity pockets are revealed. The specificity pockets (substrate binding pocket) determine the substrate specificity of the enzyme. Different pockets have distinct properties to bind to their corresponding substrates. The schematic representation of such different pockets with their substrate residues for the AVCP is shown in Figure 4.4.8.1. Inspection of different specificity pockets indicates significant conformational changes in the main-chain backbone and also in the side chain of substrate binding residues.



**Figure 4.4.8.1:** Schematic representation of different specificity pockets interacting with different substrate residues. The specificity pocket residues are written in pink color. The hydrophobic interactions are presented through dotted lines.

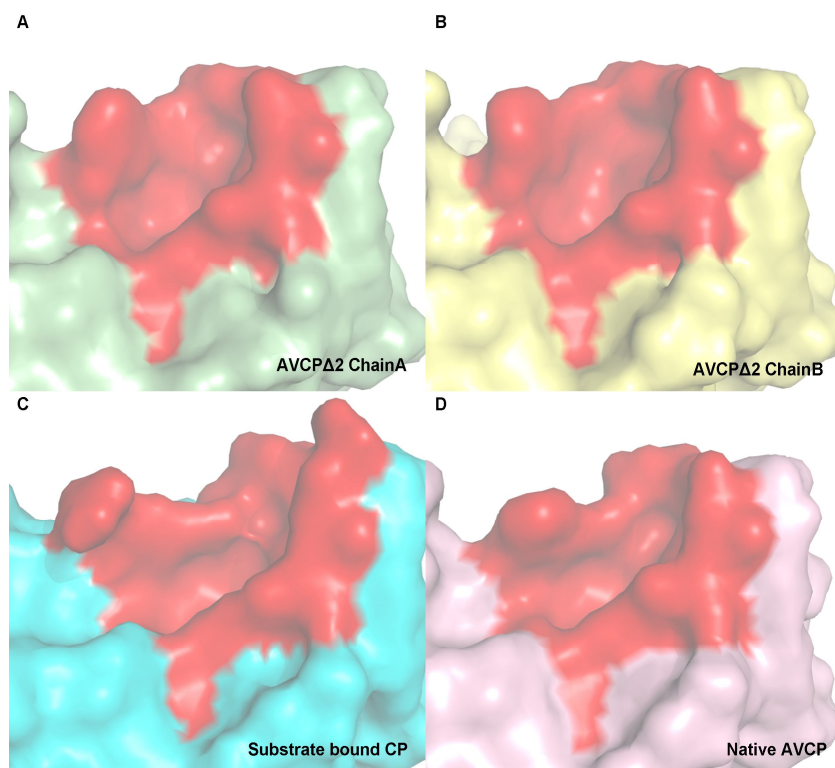
The most notable conformational changes are seen both in the backbone and amino acid side chains of the  $S_1$  specificity pocket. The  $S_1$  specificity pocket includes residues Trp194, Ile208-Ser218 (includes active site Ser residue), Ile232-Asn238 and Thr243-Leu245. The largest deviation between the active AVCP $\Delta 2$  and inactive AVCP is in the main chain of residues Ile208-Gly216. The substrate bound SCP S215A mutant representing the intermediate form, shows a conformation rather similar to the C-terminal Trp bound inactive AVCP state. This confirms the conformational flexibility of the specificity pocket between active and inactive Trp bound forms. Gly213 and Gly214 show deviation up to 2.9 Å between AVCP and Chain A of active AVCP $\Delta 2$ . The deviation is different between the two chains and found to be 3.4 Å between AVCP and Chain B of AVCP $\Delta 2$  (Figure 4.4.3.4 & 4.4.8.2). Moreover, as described above, Pro215 and Ser218 show side chain conformational flexibility in the active and inactive forms of AVCP.





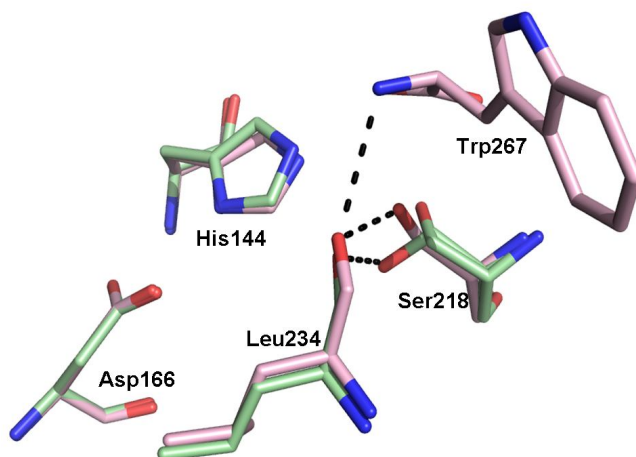
**Figure 4.4.8.2:**  $S_1$  Specificity pocket of AVCP $\Delta 2$  chain A (green) and chain B (yellow) is compared with that of native AVCP (pink). The main chain backbone at Gly213, Gly214 region is showing major difference in both chain A (2.9 Å) and chain B (3.4 Å) as compared to native form. Gly236 interacts with Val265, Asn238 with Thr210 and Leu245 with His261 in the native form which are absent in AVCP $\Delta 2$ .

Apart from some minor deviations found at region Ile232-Gly236 among the two structural forms, a displacement of up to 0.9 Å in main chain can be seen at Ala237-Thr243. This displacement may be significant, as this region lines the substrate specificity pocket. This difference can be accounted by the presence of the C-terminal residues in AVCP native structure. Due to the presence of Trp in the AVCP crystal structure bound at the  $S_1$  specificity pocket, the C-terminal residues are restricted and their interaction with neighboring residues forms a compact cavity. In AVCP $\Delta 2$ , these interactions are absent, the C-terminal loop is flexible, and does not involve in such interactions. This suggests a flexible  $S_1$  specificity pocket in the absence of Trp267. The high flexibility of the  $S_1$  pocket can be seen in figure 4.4.8.3. The surface view clearly shows the difference in the pocket between Trp bound and unbound forms.



**Figure 4.4.8.3:** Surface view of S<sub>1</sub> specificity pocket of all three enzymatic forms.

The S<sub>1</sub> specificity pocket residue Leu234 in AVCP is conserved among alphavirus CPs. Leu234 main chain carbonyl O forms an H-bond with Ser218 side chains in all the different forms of the CP. This also forms an H-bond with the NH atom of the main chain preceding Trp267 (Figure 4.4.8.4). This suggests the importance of Leu234 in holding the catalytic triad at a proper position and providing a favorable connection between the active site and substrate. The Leu234 backbone shows deviation of less than 0.4 Å between the truncated and native states of AVCP, which could be due to the absence of Trp267 as it interacts with Leu234. Ser214 (conserved in other serine proteases) which corresponds to Leu234 in alphavirus CP, is very crucial for serine protease activity as it forms an H-bond with the catalytic triad residue Asp102 (chymotrypsin residue) (Figure 4.4.8.5). Also mutational studies show a reduction in catalytic activity upon Ser214 mutation as the consequence of displacement of its backbone (Krem *et al.*, 2002). Thus, the position at Leu234 is very important for substrate binding and hence crucial for catalysis.



**Figure 4.4.8.4:** Leu234 main chain forms H-bond with Ser218 as well as Trp267.

The  $S_1$  specificity pocket determines the substrate specificity for different enzymes. A single residue present in the pocket is responsible for the recognition of specific substrate. The substrate for the trypsin enzyme consists of either lysine or arginine residue because the  $S_1$  specificity pocket of trypsin contains Asp189 residue (involved in determination of substrate specificity). The negatively charged aspartate residue shows interactions with the basic residues lysine or arginine. Likewise, the  $S_1$  specificity pocket of the chymotrypsin and chymotrypsin like serine proteases should contain small hydrophobic residue as its substrate consists of a tryptophan residue. The chymotrypsin has serine residue at this place, however the SCP and AVCP consists of valine and alanine respectively. This Ala212 of AVCP provides the hydrophobic environment as well as the ample space for the binding of tryptophan substrate to the specificity pocket.

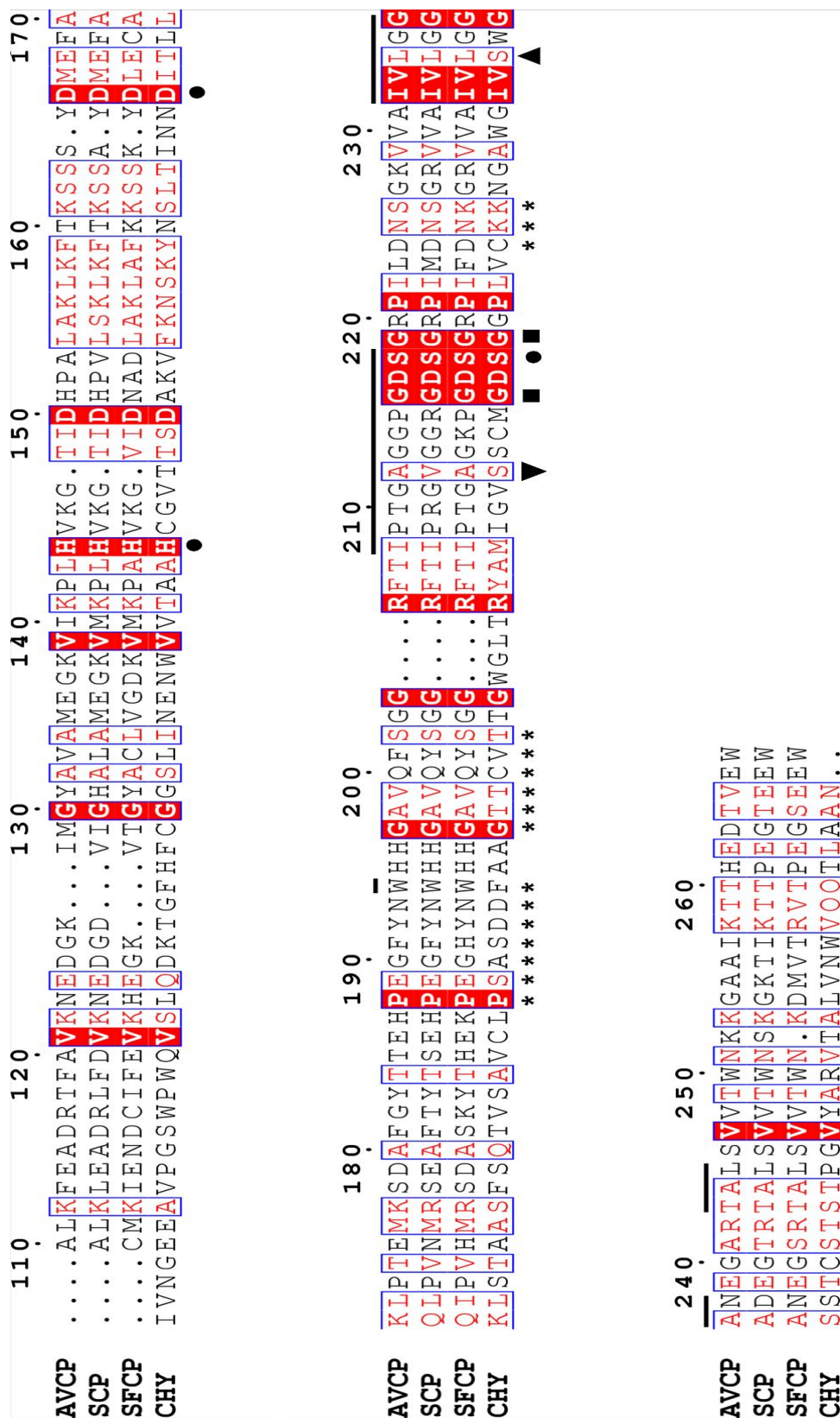
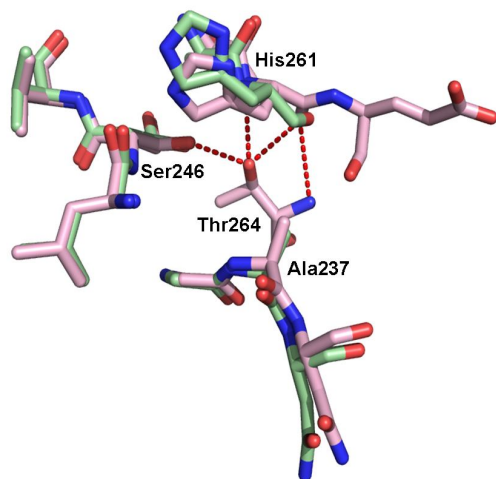


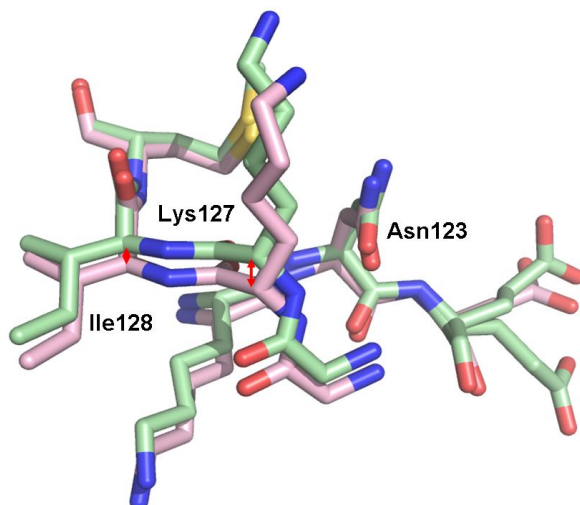
Figure 4.4.8.5: Multiple Sequence alignment of alphavirus CPs and chymotrypsin. Active site, oxyanion hole and dimer forming residues are denoted as circles, squares and stars respectively. S<sub>1</sub> pocket residues are indicated by a straight line above sequence. Leu234 and Ala212 are shown by triangle and reverse triangle respectively.

Another region of substantial flexibility lies in the S<sub>4</sub> specificity pocket that comprises of Ala237, Ser246 and His261 residues out of which Ala237 shows a backbone deviation of 0.6 Å while Ser246 shows side chain flexibility. In both forms, it is present in a dual conformation that contributes the major region of flexibility (Figure 4.4.8.6). This Ser246 side chain shows polar contacts with Leu234, Gly235, Val247, Thr259 and Thr264 in both inactive and active AVCP; one additional interaction with His261 is present in inactive form and is absent in the active AVCPΔ2. His261 also shows deviation in the main-chain backbone (0.8 Å) between the two forms of AVCP. It is probably due to the change in His261 position between different enzymatic states to bring the scissile bond near the active site. The P<sub>2</sub> and P<sub>3</sub> residues Glu266 and Val265 are present in a loop which is oriented outwards. Glu266 forms ionic interaction with catalytic triad residue His144. However, in the active AVCPΔ2 crystal structure Val265 is found to be disordered.



**Figure 4.4.8.6:** Comparison of the S<sub>4</sub> pocket shows deviation in backbone as well as side chains. Ser246 interaction with His261 and Thr264 is found in inactive AVCP, but absent from the active one.

Likewise, the S<sub>1</sub>' pocket (including residues Met129, His144 and Val145) does not show much variation (~ 0.4 Å) in the backbone conformation. Ser265, the P<sub>1</sub>' residue in SCP substrate bound form shows H-bonding with Ile126, which corresponds to Met129 in AVCP. Met129 shows polar interaction with the main chain of Val121. Again the significant variations lie in the S<sub>4</sub>' and S<sub>2</sub>' specificity pockets. The S<sub>4</sub>' pocket residues Asn123 and Lys127 are displaced from the position by 0.5 Å and 0.9 Å, respectively. This pocket consists of Asn120 and Asp124 in SCP, however the P<sub>4</sub>' residue is hydrophobic in both (Ile271 in AVCP and Pro268 in SCP). Ile128 and Met129 residues in S<sub>2</sub>' pocket shows significant deviation of 0.6 Å and 0.4 Å respectively (Figure 4.4.8.7). The P<sub>2</sub>' residue Ala266 in the SCP substrate bound form shows interaction with water molecules.



**Figure 4.4.8.7:**  $S_4'$  and  $S_2'$  pockets comparison in active and inactive AVCP demonstrates the displacement of 0.5, 0.9 and 0.6 Å in Asn123, Lys127 and Ile128 respectively.

#### 4.5 Conclusion

In last two decades extensive investigation on the alphavirus CP has been performed. Most of the structure-function analyses have focused on characterization of the inactive form of CP (C-terminal Trp bound form) using X-ray crystallographic, cryo-EM, mutational and biochemical studies (Tang *et al.*, 2011; Thomas *et al.*, 2010; Hahn *et al.*, 1985). In 2008, the enzymatic characterization of the esterase activity of the truncated form of SFV CP was reported (Morillas *et al.*, 2008). It was proposed that the conserved Trp residue at the C-terminus is required for proper folding and stabilization of CP and the active capsid produced by truncating the conserved C-terminal Trp is natively an unfolded protein. As a result, the three dimensional structure of the active form of CP has not been determined.

As a consequence, the protease activity of alphavirus CP has not been targeted for antiviral drug discovery. In this study, recombinant active AVCP in which two C-terminal  $P_2$  and  $P_1$  residues Glu266 and Trp267 have been removed (AVCP $\Delta 2$ ), was produced and purified. The crystal structure of this active form of the Aura virus CP (AVCP $\Delta 2$ ) has been determined. This is the active pre-cleavage state preceding cis-catalytic activity. Thus, there are three different forms of the alphavirus CP: the active state (this chapter), the substrate bound intermediate form (Choi *et al.*, 1996) and the inactive C-terminal Trp-bound form (Chapter 2). It is likely that the protein undergoes conformational rearrangements during structural polyprotein processing, viral replication and budding. The significant conformational changes in the three different forms have been analyzed and reported in this study. The novel structure of the active form of AVCP $\Delta 2$

provides structural insight into the catalytic site and substrate binding pockets revealing a range of conformational flexibilities. These conformational variations should be taken in account for structure based drug designing of anti-alphaviral molecules targeting CP activity.

Structural comparisons show that the specificity site pockets are highly flexible. The active site residue Ser218 side chain is in a dynamic state and exhibits different conformations. The Leu234 backbone has an important role in substrate binding and maintaining the position of the catalytic triad. The oxyanion hole contains a water molecule in the truncated form and Chain A of substrate bound form; however, it is missing from Chain B of the substrate bound form and native enzymatic state. This might be due to the formation of an H-bond between the scissile bond and the oxyanion hole residues during proteolysis. Moreover, the last C-terminal segment is highly flexible until the substrate attaches to the pocket and undergoes proteolysis. This loop is missing in the unblocked form, flexible in the substrate bound form and static in C-terminal Trp-bound native state. Knowledge of the structure of the truncated unblocked alphavirus CP allows for a comprehensive analysis of capsid structures at different stages. The presence of glycerol in the active site of AVCP $\Delta$ 2 where the P<sub>1</sub> residue Trp267 binds opens the possibility of designing glycerol-based alphavirus CP inhibitors. The reported structure highlights the previously unseen conformational changes and provides a foundation for structure-based design of antiviral compounds that will block the initial step of alphavirus structural polyprotein processing.





## 5.1 Abstract

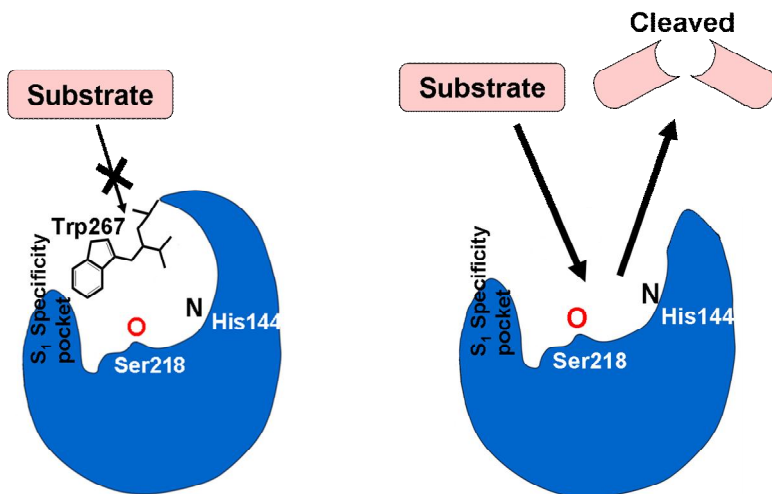
The protease activity of the alphavirus capsid is a potential anti-alphaviral target to arrest capsid release, maturation and structural polyprotein processing. However, the discovery of the capsid protease inhibitors has been hindered due to the lack of a suitable screening assay. Chikungunya virus (CHIKV) is another member of alphavirus genus which is an arthropod-borne, highly pathogenic virus that causes a large number of human infections each year. The 2006 outbreak of CHIKV leads to the reemergence of the virus responsible for Chikungunya fever, joint pains and persistent arthralgia. As a public health concern, its prevention and treatment has become a necessity, but unluckily no vaccine or antiviral is existing that can prevent or treat CHIKV infection. Thus, it is important to formulate a strategy for the development of antivirals against CHIKV. This chapter demonstrates the trans-proteolytic activity of CHIKV capsid protein (CVCP) as well as AVCP which can be used as high throughput compound screening assay against alphavirus infection.

Being the initial step in structural polyprotein processing, the proteolytic activity of CVCP is a potential drug target against CHIKV. In this Chapter, we report the development of a trans-proteolytic activity assay for CVCP based on fluorescence resonance energy transfer (FRET) for screening protease inhibitors. The  $K_M$  and  $K_{cat}/K_M$  values were determined using fluorogenic substrate and were found to be  $1.27 \pm 0.34 \mu\text{M}$  and  $5.5 \times 10^4 \text{ M}^{-1} \text{ min}^{-1}$  respectively. The effect of buffer pH, ionic strength, glycerol concentration and the increasing enzyme concentrations on the proteolytic activity of CVCP has been investigated. The kinetic parameters are almost similar to that of the AVCP. Kinetic parameters using fluorogenic peptide substrates were also estimated for AVCP, and a  $K_M$  value was found to be  $2.63 \pm 0.62 \mu\text{M}$  and a  $K_{cat}/K_M$  value was  $4.97 \times 10^4 \text{ M}^{-1} \text{ min}^{-1}$ .

In this study, we have described the formerly unappreciated *trans*-proteolytic activity of the enzyme and for the first time have developed a FRET based protease assay for the screening of alphavirus capsid protease inhibitors. The availability of active recombinant CVCP enzyme in combination with the fluorogenic peptide based *in vitro* FRET assay can be used for screening and development of CHIKV capsid protease inhibitor molecules. Likewise, this assay can be used for capsid protease from other members of the alphavirus genus leading to the development of anti-alphaviral compounds.

## 5.2 Introduction

The capsid protein (CP) is reported to not possess trans-activity since it is an auto-proteolytic enzyme that auto-inhibits to give a turn-over number of 1 (Choi *et al.*, 1991; Hahn & Strauss, 1990). However, investigations by Morillas *et al.* have revealed that truncations of 1-7 C-terminal residues of the SFV CP, including deletion of the ultimate C-terminal tryptophan, restores the enzymatic activity of protein (Morillas *et al.*, 2008) (Figure 5.2.1). The truncated enzyme was found to possess high esterase activity after deletion of the highly conserved C-terminal tryptophan. However, until now the trans-cleavage protease activity of the alphavirus CP in which trans-cleavage of the peptide bond containing tryptophan at P<sub>1</sub> position has not been demonstrated. Furthermore, they concluded from the study that correct folding of the tertiary structure of SFV CP is dependent on the presence of conserved Trp residue at the C-terminus (Morillas *et al.*, 2008). These interesting findings of the study persuaded us to probe the trans-peptidase activity of another member of alphavirus genus, AVCP and CVCP.



**Figure 5.2.1: Schematic representation of inactive and active forms of alphavirus CP.** In the inactive form Trp267 shows autoinhibitory action on enzyme activity after autoproteolysis (left). In the absence of Trp267, the protein acquires its catalytic property as the substrate can easily access the active site (right).

Chikungunya virus infection has been characterized by high fever, muscle pain, gastrointestinal troubles, headache, rashes, eye and neurological problems. It is arthropod-borne virus and transmitted through mosquitoes *Aedes aegypti* and *Aedes albopictus*. The incubation period for the Chikungunya fever ranges from 2-12 days followed by an acute phase that lasts for days to weeks. In some cases, severe Chikungunya infection leads to chronic disease characterized by persistent arthralgia from weeks to years (Larrieu *et al.*, 2010). CHIKV was first isolated in 1953 in Tanzania (Ross, 1956) and afterwards it became an epidemic in Asia and Africa from the

period 1954 to 2000 (Powers & Logue, 2007; Ligon, 2006). Subsequently, major outbreak started in 2004 with an epidemic in Kenya (Sergon *et al.*, 2004) that leads to the spread of CHIKV in Madagascar, Comoro, Mayonette and La Réunion islands of Indian Ocean, India, West Africa and South-East Asia (Sergon *et al.*, 2007; Padbidri & Gnaneswar, 1979; Nimmannitya *et al.*, 1969; Shah *et al.*, 1964). According of the data provided by Centers for Disease Control and Prevention (CDC) in 2014, more than 60 countries and territories are listed from where Chikungunya cases have been reported (<http://www.cdc.gov/chikungunya/geo/index.html>). CDC data also show the CHIKV infection in the islands of Caribbean for the first time in late 2013. Hence, it is emerging day by day and arise the need of a proper measure (vaccine and therapeutics) against its infection.

Besides Chikungunya fever, the infection also includes lymphopenia, lethal hepatitis, encephalitis and neonatal encephalopathy (Robin *et al.*, 2008; Lenglet *et al.*, 2006). Due to A226V mutation in E1 glycoprotein (involved in viral fusion) of CHIKV, an additional mosquito vector *Aedes albopictus* was observed during the La Réunion epidemic (Schuffenecker *et al.*, 2006). Thus, the high pathogenicity and the increase in the disease transmitting vector makes CHIKV crisis more intense and needs immediate attention towards its eradication or virus specific treatment. Recently, a number of attempts have been made to develop vaccines or antivirals against CHIKV infection. The virus like particles (VLPs) of Chikungunya virus has been proposed to be used as vaccines against CHIKV (Akahata *et al.*, 2010). Also, the immunoglobulins transfer and IFN- $\alpha$  treatment have provided some hint for the prevention of CHIKV infection (Gardner *et al.*, 2010; Couderc *et al.*, 2009). These are not cost-effective and can only be used before the virus infection; hence the alternative methods are required to be developed. Despite of the vaccine preparation trials, a number of antivirals such as ribavirin (a broad spectrum antiviral) and Chloroquine (considered as to block virus entry and maturation) were tested for antiviral activity *in vivo* with no positive results (De Lamballerie *et al.*, 2008; Ozden *et al.*, 2008; Ravichandran & Manian, 2008; Briolant *et al.*, 2004). However, Chloroquine was found to have beneficial effect on chronic arthralgia due to the infection of CHIKV (Brighton, 1984). Hence, there is an urgent need of effective and economical antiviral therapy against CHIKV infection.

Chikungunya virus belongs to the alphavirus genus of *Togaviridae* family having a positive sense single stranded RNA genome. The structural polyprotein is translated from the 26S subgenomic RNA (Strauss & Strauss, 1994). The CP is present at the N-terminus and consists of two domains: the amino terminal and the carboxyl terminal domains. The amino-terminal domain is involved in capsid-capsid interaction to form nucleocapsid core assembly (Hong *et al.*, 2006;

## Chapter 5: trans-Proteolytic activity of Alphavirus capsid protease

Forsell *et al.*, 1995). The carboxyl terminal domain of CP acts as serine protease which undergoes autoproteolysis to separate itself from rest of the polyprotein (Choi *et al.*, 1997; Tong *et al.*, 1993; Choi *et al.*, 1991; Melancon & Garoff, 1987).

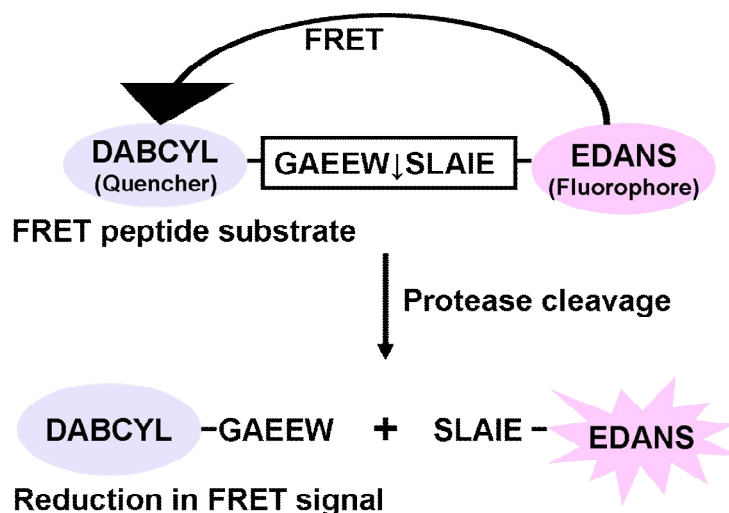
The CP is a chymotrypsin like serine protease and consists of the catalytic triad residues Ser, His and Asp similar to other serine proteases (Hahn & Strauss, 1990). The GDSG motif which consists of the active site residue Ser is conserved in all the serine proteases including CVCP. The cleavage site of CP includes the conserved C-terminal Trp-Ser between which it gets cleaved. After detaching from the polyprotein, the carboxyl-terminal tryptophan remains bind to the active site and thus makes the protein further inactive (Skoging & Liljestrom, 1998; Choi *et al.*, 1996; 1991). The binding of tryptophan in the specificity pocket makes it inaccessible for other substrates to bind the pocket.

The residue Ile227 in CVCP is present near the active site. The error prone PCR generated the I227K mutation in the CP. This mutant construct was unable to show the protease activity. This loss in activity might be due to the formation of salt bridge between Lys227 and Asp161 (catalytic triad residue) which result in the disruption of the electrostatic interactions between the catalytic triad residues. Thus, in CVCP, a single mutation makes the enzyme inactive (Thomas *et al.*, 2010).

Capsid protease represents very significant targets for anti-CHIKV drug development as involved in the very first step in structural polyprotein processing. Fluorogenic substrates are used for the high throughput compound screening of a number of viral proteases such as SARS coronavirus 3CLpro protease, West Nile virus Serine Protease, HIV (Human immunodeficiency virus) protease, Dengue NS2B-NS3 protease, HCV (hepatitis C virus) NS3/4A protease (Nitsche *et al.*, 2013; Jin *et al.*, 2011; Sabariego *et al.*, 2009; Mueller *et al.*, 2008; Lin *et al.*, 2004). This technique is very simple and highly sensitive, hence provides a tool for high throughput screening of antivirals.

The characterization of the trans-proteolytic activity of the CVCP and the screening of inhibitors are the main aim of this study. As the protease activity of CVCP is the first step in structural polyprotein processing, hence, it can be a potential drug target against CHIKV infection. The proteolytic activity of the CVCP has been monitored with the use of 4-(4-dimethylaminophenyl-azo)benzoic acid (DABCYL) and 5-[(2-aminoethyl)amino]naphthalene-1-sulfonic acid (EDANS) on the sides of the peptide which contains the target sequence (Figure 5.2.3). This pair demonstrates FRET (fluorescence resonance energy transfer) in which EDANS acts as fluorophore and DABCYL as quencher which quenches the fluorescence produced by

EDANS. After getting cleaved by the protease, these two gets separated and the FRET signal is reduced. As a result, the increase in the fluorescence takes place which can be monitored to measure the enzyme activity. Using a fluorescence plate reader, this assay can be easily performed in a 96- and 384- well plate format for high throughput screening of inhibitor molecules against the CHIKV capsid protease. A high throughput protease assay is cost effective as it is performed in a small volume with the use of lesser enzyme as well as substrate and also reduces the experimental time for inhibitor screening.



**Figure 5.2.2: Fluorescence resonance energy transfer (FRET) assay.** A FRET based protease assay has been developed for the activity determination of alphavirus CP. The substrate is designed on the basis of the sequence of the peptide from the C-terminus of the protein where it gets recognized and cleaved by the protease. The DABCYL (Quencher) and EDANS (fluorophore) are attached at the N and C terminus respectively. The donor fluorophore EDANS produces the signal upon excitation which gets quenched by the quencher DABCYL and thus shows FRET. In the presence of the protease, the cleavage takes place and the donor fluorophore and quencher gets separated from each other which results in the reduction of FRET signal. This further leads to the increase in the fluorescence signal of the donor fluorophore. In the presence of protease inhibitors, the FRET should remain same and no any increase in the fluorescence should occur.

To fight against Chikungunya infection either vaccine or drug development is highly essential. However, only a few studies are available those are involved in the inhibition of virus replication. HIV protease inhibitors are well established and approved by Food and Drug Administrations (FDA) (<http://www.fda.gov/ForConsumers/byAudience/ForPatientAdvocates/HIVandAIDSActivities/ucm118915.htm>). Likewise, Chikungunya capsid protease, involved in the first step of structural polyprotein processing, can also be considered a potential target to develop antivirals against CHIKV infection. Moreover, previous studies include the cell based assays that require either BSL3 or BSL2 (Biosafety Level 3 or 2) containment (Deu John *et al.*, 2013). However, our *in vitro* assay is very simple and does not need any BSL2 or BSL3 containment.

## Chapter 5: trans-Proteolytic activity of Alphavirus capsid protease

In this Chapter, we have successfully cloned the gene encoding CVCP through molecular cloning of cDNA in expression vector, expressed the soluble form of protein in *E. coli* and purified the protein to homogeneity. The AVCP and CVCP both have been characterized for *in vitro* proteolytic activity which was performed using highly sensitive fluorogenic peptide substrate. Additionally, the kinetic parameters for CVCP trans-protease activity were calculated using the FRET based assay. A known protease inhibitor PMSF (phenylmethylsulfonyl fluoride) and few tryptophan based inhibitors were tested for the inhibition of trans-proteolytic activity of CVCP. The gel filtration studies were performed to get the knowledge of oligomeric state of CVCP.

### 5.3 Materials and methods

#### 5.3.1 Molecular cloning and construction of expression plasmid

The CHIKV genomic RNA (procured from DRDE, Gwalior) was denatured at 70 °C for 5 minutes and chilled on ice. The first strand cDNA was synthesized using oligo(dT)<sub>30</sub> primer with M-MulV reverse transcriptase at 37 °C for 1 h. The reaction was carried out in 20 µl of reaction volume and treated with RNasin. The primers for recombinant DNA cloning were designed according to the sequence provided in NCBI (National Center for Biotechnology Information) (Genbank accession no. AEA10291.1 for structural polyprotein and ADZ47899.1 for capsid protein) (Table 5.1).

CHIKV DNA fragment (~ 3.7 kbp) encoding the structural polyprotein was amplified using the synthesized cDNA as a template. The oligonucleotides F1 and R1 were used for the amplification of structural polyprotein. The amplified product was again subjected to polymerase chain reaction (PCR) with the primers F2 and R2 having *Nde*I and *Bam*HI restriction enzymes sites. The PCR product was purified with the use of PCR purification kit from Qiagen and restriction digestion was done for the purified PCR product. The pET28c vector was also digested using the same restriction enzymes. Both the digested products were subjected to agarose gel electrophoresis and the digested bands were excised. The DNA was eluted from these excised bands of agarose gel using DNA gel extraction kit (Qiagen, USA). The two DNA fragments were ligated overnight at 15 °C with T4 DNA ligase. The ligated mixture was transformed into *E. coli* cells *DH5*α (DE3) and the culture was spread on Luria Bertani (LB) agar plate having 50 µg/ml kanamycin. The plate was incubated overnight at 37 °C. The colonies were picked and screened for the presence of structural polyprotein encoding gene insert.

The confirmed constructs containing the DNA encoding structural polyprotein of CHIKV was used as a template for the amplification of both the native, CVCP (residues 106-261) and the C-terminal truncated CVCP $\Delta$ 2 (residues 106-259). The amplification and cloning of CP was carried out using oligonucleotides, F3, R3 (for inactive CVCP) and R4 (for active CVCP) having *Nde*I and *Xho*I sites. The PCR amplified products were subcloned into pET28c vector using the same protocol as described above. Positive clones containing the insert of correct size were identified by doing PCR and restriction enzyme digestion of isolated plasmids. Further, DNA sequencing of the isolated plasmid was done to confirm the presence of DNA fragments encoding CVCP and truncated CVCP $\Delta$ 2 in the vectors. To facilitate protein purification using Ni<sup>+2</sup> affinity chromatography, the N-terminal Tobacco Etch Virus (TEV) protease cleavable His<sub>6</sub>-tag was incorporated in both the constructs.

**Table 5.1:** List of the oligonucleotides used

Name	Sequence
F1	5'- ATGGAGTTCATCCCAACCCAAACTTTTTACAATAGG AGG -3'
R1	5'- TTAGTGCCTGCTGAACGACACGCATAGCAC -3'
F2	5'- ACGAACATATGATGGAGTTCATCCCAACCCAAAC -3'
R2	5'- AAGCAGGATCCTTAGTGCCTGCTGAACGACAC -3'
F3	5'- CTGGAATTCATATGTGCATGAAAATCGAAAATGATTGTATTTTCG -3'
R3	5'- CTAGAATCTCGAGCTACCA CTC TTC GGC CCC CTC -3'
R4	5'- CTA GAA TCT CGA GCT ATT CGG CCC CCT CGG G -3'

### 5.3.2 Expression of both inactive and active CVCP

The expression of cloned construct was optimized to get the good amount of soluble protein. Firstly, both the recombinant plasmid constructs were transformed in *E. coli* strain *Rosetta* (DE3) and plated on LB agar plates containing kanamycin (50  $\mu$ g/ml) and chloramphenicol (35  $\mu$ g/ml). The plate was incubated overnight at 37 °C and a single colony was picked to grow in LB broth supplemented with the same antibiotics. The secondary culture was grown using inoculum from the overnight primary culture at 37 °C till the OD at 600 nm reaches

## Chapter 5: trans-Proteolytic activity of Alphavirus capsid protease

to 0.5 (OD<sub>600</sub>). Afterwards, the culture for inactive CVCP was induced with 0.4 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) when it attains the OD of 0.8 and allowed to grow at 37 °C for 4 h. For the expression of active CVCP, the culture was transferred to 18 °C till OD<sub>600</sub> of 0.7 and induced using 0.4 mM IPTG. The culture was further grown at 18 °C overnight. The cells for both the proteins were harvested by centrifugation and analyzed for the expression of soluble proteins using 15 % SDS-PAGE.

### 5.3.3 Purification of CVCP (inactive and active)

The purification of both the proteins was performed using the pellet from 1 liter culture. The pellet was re-suspended in 25 ml of binding buffer (50 mM Tris-HCl pH 7.6, 15 mM imidazole, and 100 mM NaCl) on ice. The re-suspended cells were disrupted using the cell disruptor (Constant Systems Ltd, Daventry, England) and centrifuged at 4 °C to separate the supernatant and pellet. The Ni-NTA beads (pre-equilibrated with the binding buffer) were loaded with the clarified supernatant and incubation was done for half an hour at 4 °C. The His tagged purified active and inactive CVCP were eluted with 250 mM Imidazole. A HiLoad Superdex 75 16/60 size-exclusion chromatography column (GE Healthcare) was equilibrated with the buffer (50 mM Tris-HCl pH 7.6, 20 mM NaCl). The protein eluted from the affinity column was concentrated using a 3 kDa cutoff Amicon Ultra-15 concentrator (Millipore, Bedford, Massachusetts, USA) and loaded onto the gel filtration column. ÄKTA purifier (GE Healthcare) at 4 °C was used to run gel filtration column at a flow rate of 0.5 ml/min. The major peak fractions were collected and analyzed using SDS-PAGE. The purified fractions were pooled and concentrated again. The molecular weight markers were also run on size exclusion column for the determination of the molecular weight of the protein and preparation of standard curve. The concentration of the purified and concentrated protein was measured by UV-Vis spectroscopy at a wavelength of 280 nm using an extinction coefficient method.

### 5.3.4 trans-Protease activity assay

The proteolytic activity assay for active AVCP and CVCP was performed by a FRET-based proteolytic assay using 5-[(2'-aminoethyl)-amino] naphthalenesulfonic acid (EDANS) and 4-[[4-(dimethylamino) phenyl] azo] benzoic acid (DABCYL) as the FRET pair. The peptide having DABCYL and EDANS at the N and C termini, respectively, was procured from Biolinkk, New Delhi, India. The sequence of the substrate peptide was derived from the cleavage site of CP



containing the conserved Trp-Ser scissile bond for protease cleavage. The cleavage analysis was performed in 20 mM HEPES buffer (pH 7.0) (reaction buffer) at room temperature. The purified enzyme was incubated with the reaction buffer and the fluorogenic peptide substrate was added to a final concentration of 1  $\mu$ M. The reaction was carried out in total volume of 1 ml in an eppendorf tube. At various time points, the reaction mixture was transferred to the fluorescence cuvette for steady state fluorescence measurements with a Fluorolog-3 Spectrofluorimeter LS55 (HORIBA Jobin Yvon Spex). The fluorogenic peptide substrate was excited at a wavelength of 340 nm and the emission spectrum scanning from 440 to 600 nm was recorded. The increase in the fluorescence was monitored at specific time intervals. For kinetic studies, different peptide substrate concentrations ranging from 0.6 to 16  $\mu$ M were used. The initial velocity ( $V_i$ ) at different substrate concentrations was calculated. The kinetic parameters were calculated by Lineweaver Burk Plots. The readings were taken after specific time intervals accompanied with the control reaction having no enzyme. All the readings were normalized using the control reaction. The enzymatic reaction was also performed with the use of different concentrations of enzyme. To see the effect of enzyme concentration, an increasing concentration of the enzyme (100 nM to 800 nM) was used in the reaction and the activity assay was performed following the above protocol. The tryptophan based inhibitors and basic protease inhibitors were screened for the inhibitory effect on the CVCP activity. For the inhibition studies, the compounds were pre-incubated with the protein for 2 hrs and the assay was performed using the same protocol described above.

#### **5.3.5 Effect of pH and NaCl concentration on enzyme activity**

The effect of buffer pH on enzymatic activity was observed using different buffers ranging from pH 4.5 to 9.5. The buffers used were 20 mM sodium acetate (pH 4.5, 5.0, 5.5), MES (pH 6.0, 6.5), HEPES (pH 7.0, 7.5), Tris-HCl (pH 8.0, 8.5), Bicine (pH 9.0) and Glycine-NaOH (pH 9.5). The initial velocity for the enzyme in all the buffers was calculated and the relative activity was measured. The influence of NaCl on protease activity was analyzed by performing the reaction in the presence of different concentrations of NaCl ranging from 0 to 500 mM.

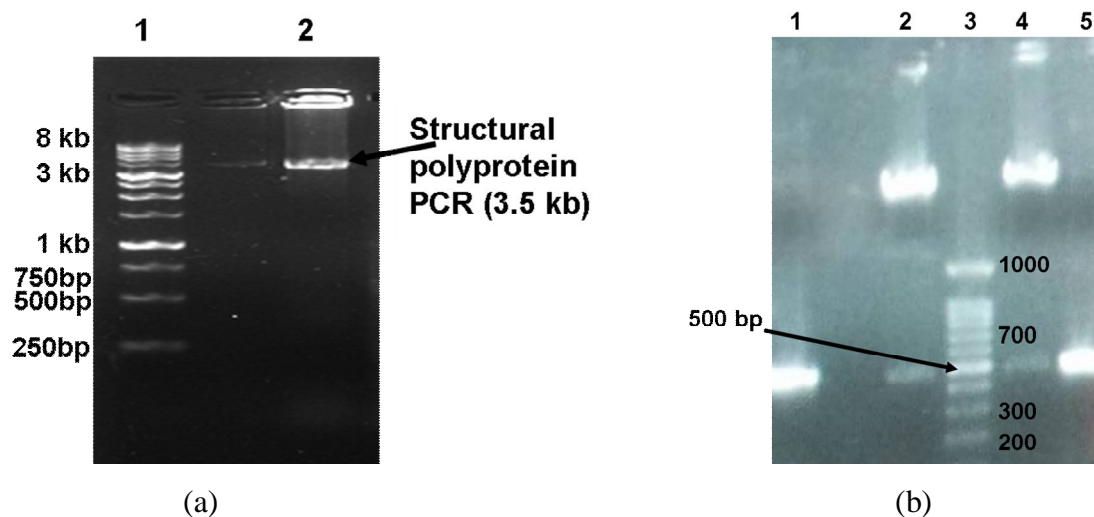
#### **4.3.6 Effect of glycerol**

The effect of increasing glycerol concentration was also observed using different concentrations of glycerol in the reaction buffer. The concentration of the glycerol used was ranging from 10 % to 50 %. The activity assay was carried out as above.

## 5.4 Results and Discussion

### 5.4.1 Expression and Purification of recombinant CVCP (both active and inactive forms) in *E. coli*

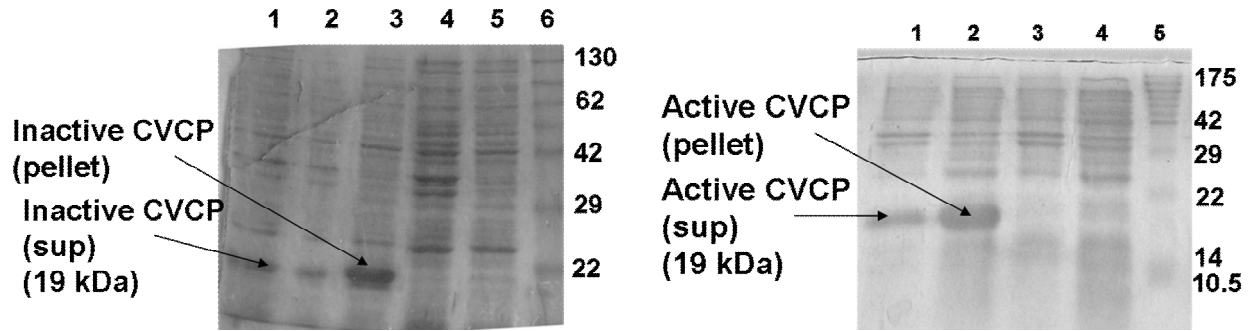
After autoproteolysis and the release of CP from structural polyprotein, Trp261 remains bound near the active site and makes the protein further inactive for either cis or trans autoproteolytic action. However, previous studies on alphavirus CP indicate its trans activity upon removal of this C-terminal tryptophan (Morillas *et al.*, 2008). So, in order to analyze the enzymatic activity of CVCP, both inactive and the active proteins were expressed and purified. The structural polyprotein was amplified using the cDNA of CHIKV genome (Figure 5.4.1.1a) and cloned into the vector. This plasmid construct having the structural polyprotein encoding gene was used for further experiments. The protease domain of CVCP was amplified from the construct with structural polyprotein encoding gene. This amplified product was further cloned into pET28c vector to form the expression plasmid pET28c-CVCP (Figure 5.4.1.1b).



**Figure 5.4.1.1a:** PCR amplification of structural polyprotein from Chikungunya virus; **b:** Cloning of inactive and active CVCP. Lane 1, PCR Amplified active (110-259) protease gene; Lane 2, RE digested active (110-259) protease gene; Lane 4, RE digested inactive (110-261) protease gene; Lane 5, PCR Amplified inactive (110-261) protease gene; Lane 3, DNA Ladder.

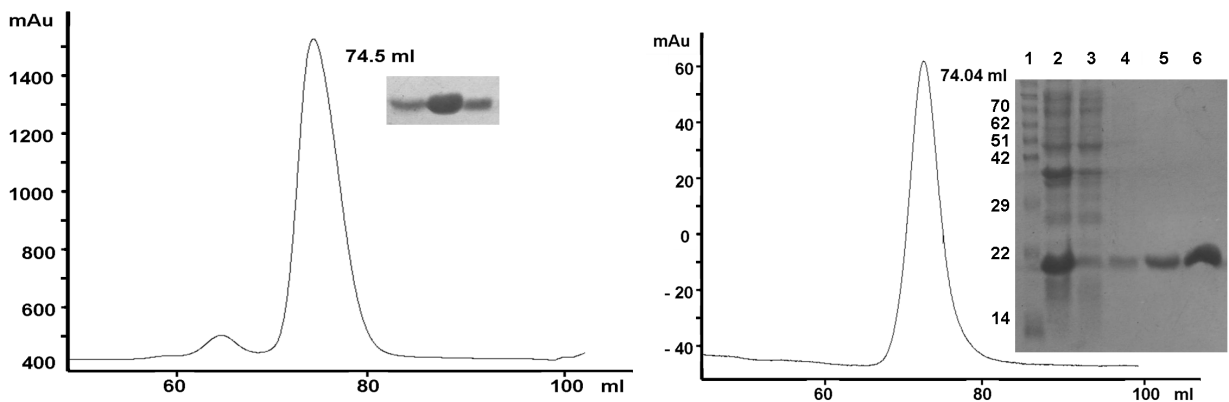
The inactive protein was containing the full protease domain (106-261 residues). Truncations for the last two residues including C-terminal Trp261 from the inactive construct results in the generation of active form of CVCP (106-259 residues). Both the constructs were cloned successfully into expression vector.

The optimization of the expression conditions for both the proteins was done to get the good amount of soluble protein. After optimization of induction temperature, induction time and IPTG concentration, both the proteins having the 6X-His tag at the N-terminus were expressed in the soluble form (Figure 5.4.1.2).



**Figure 5.4.1.2:** SDS-PAGE analysis of expression and solubility of recombinant Chikungunya inactive (110-261) CP (left panel). Lane 1 & 2, Induced pellet; 3, Induced supernatant; 4, Uninduced pellet; 5, Uninduced supernatant; 6, Protein marker. SDS-PAGE analysis of expression of Chikungunya active (110-259) CP (right panel). Lane 1, Induced pellet; 2, Induced supernatant; 3, Uninduced pellet; 4, Uninduced supernatant; 5, Protein marker.

The soluble proteins were purified by IMAC (immobilized metal assisted chromatography) and size exclusion chromatography. In Ni-NTA column chromatography, the protein was eluted by increasing concentration of imidazole. A single band was observed on SDS-PAGE for both the purified proteins confirming the protein homogeneity (Figure 5.4.1.3). The protein band was visible at ~ 19 kDa on SDS gel which was also confirmed by gel filtration chromatography by comparing and calculating the molecular weight using standard molecular weight marker. Thus, it is confirmed that both inactive and active forms of CVCP are monomer in solution.

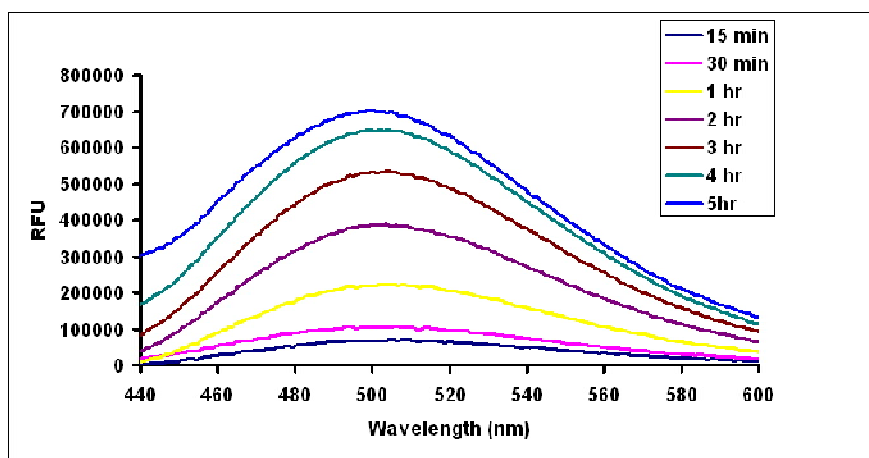


**Figure 5.4.1.3:** SDS-PAGE and gel filtration profile of both inactive (left panel) and active (right panel) CVCP shows the protein purified to homogeneity. The size-exclusion chromatography results suggest the monomeric nature of both the proteins. Lane 1, molecular-weight markers (kDa); lane 2, pellet containing insoluble protein fraction; lane 3, supernatant containing soluble protein fraction; lane 4-6, purified His-tagged CVCP.

### 5.4.2 *In vitro* trans-protease assay of CVCP

The trans-proteolytic assay was performed using highly sensitive FRET based peptide substrate. The peptide sequence was designed on the basis of the amino acid sequence at the C-terminus of capsid and N-terminus of E3 glycoprotein of CHIKV which includes the residues forming scissile bond. The fluorogenic peptide consists of EDANS and DABCYL as fluorophore and quencher respectively. The cleavage of the peptide by the active CVCP separates EDANS and DABCYL which results in the increase in the fluorescence signal that was quenched previously due to the presence of DABCYL. The excitation wavelength of EDANS is 340 nm which give emission spectrum at 490 nm. This emission fluorescence intensity is increased when the quencher DABCYL gets separated by the proteolytic action of active CVCP (Figure 5.2.2).

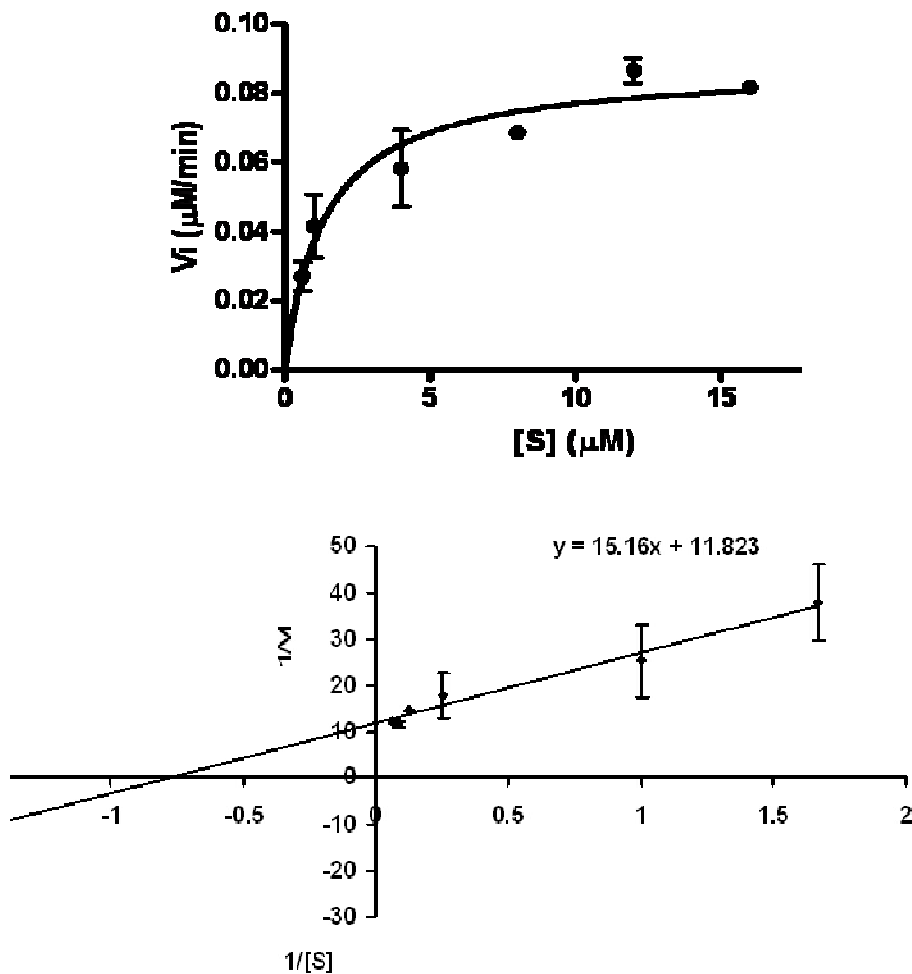
The proteolytic reaction was carried out in 20 mM HEPES buffer (pH 7.0). The enzyme was incubated in the buffer before adding the substrate. The substrate was added to the reaction and the fluorescence was measured at different time points. The increase in fluorescence was recorded with time (Figure 5.4.2.1). The RFU (relative fluorescence unit) for all the time points was calculated by subtracting the fluorescence readings of the control from the fluorescence readings of the reaction. The figure 5.4.2.1 clearly shows the increase in the fluorescence with increasing time period.



**Figure 5.4.2.1:** *In vitro* trans proteolytic assay for CVCP. The assay was performed in 20 mM HEPES buffer with 20  $\mu$ g of enzyme. The substrate was added to the final concentration of 8  $\mu$ M to initiate the proteolytic reaction and the fluorescence was measured at different time points. The fluorescence was measured till 5 hrs and the increase in fluorescence was observed. All the values are the average of triplicate data. The values were normalized using the reaction performed in the similar conditions with no enzyme.

According to the data for increase in fluorescence, the relative fluorescence units were calculated and  $V_i$  of the enzyme at different substrate concentrations were determined (Figure

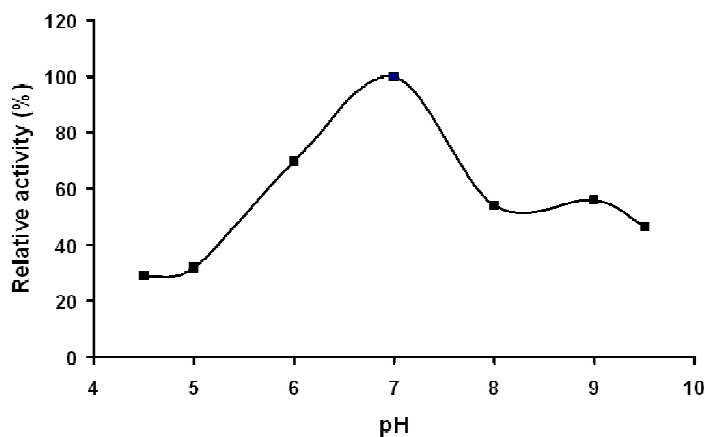
5.4.2.2a). All the values were normalized by subtracted the readings obtained with the same reaction having no enzyme. The kinetic parameters  $K_M$  and  $K_{cat}/K_M$  were calculated to be  $1.27 \pm 0.34 \mu\text{M}$  and  $5.5 \times 10^4 \text{ M}^{-1} \text{ min}^{-1}$  respectively (Figure 5.4.2.2b). The inactive CVCP was also used in proteolytic activity assay as a negative control and as expected, this construct shows the insignificant difference in fluorescence readings measured with respect to time. The calculations for the initial velocity were performed using the extinction coefficients at all the different substrate concentrations.



**Figure 5.4.2.2: Kinetic studies of CVCP. a:** To perform the kinetic studies of active CVCP, different concentrations of the substrate ranging from  $0.6 \mu\text{M}$  to  $16 \mu\text{M}$  were used and the values of  $V_{max}$  and  $K_M$  were determined. The extinction coefficient was determined for all the substrate concentrations and initial velocity was calculated. The kinetic data were fitted into the Michealis-Menten equation. **b:** The lineweaver-burk plot was created by taking the reciprocal of  $V_i$  and substrate concentrations. The intercept and the slope were calculated according to the equation  $y = mx + c$ . From these, the values of  $V_{max}$  and  $K_M$  were determined and the  $K_{cat}$  was also calculated by dividing  $V_{max}$  with the enzyme concentration.

### 5.4.3 Effect of buffer pH on enzyme activity

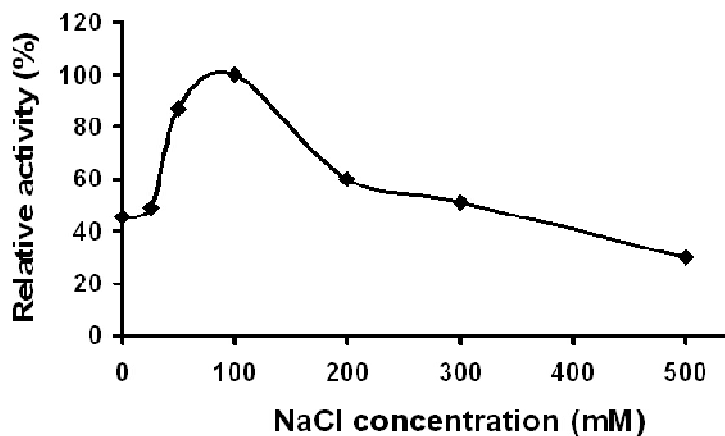
For the optimization of the reaction conditions for *trans*-proteolytic activity of active CVCP, different buffers in pH ranges from 4.5 - 9.5 were used in the reaction. The  $V_i$  was calculated at all different pHs and plotted against the buffer pH. The graph shows a regular pattern of enzymatic reaction which is optimum at pH 7.0. Above and below this pH, it shows relatively lower enzyme activity (Figure 5.4.3.1). Thus, the *trans*-proteolytic reaction favors at physiological pH 7.0 which might mimic the environment inside the mammalian cell where it undergoes *cis*-proteolytic cleavage.



**Figure 5.4.3.1:** The effect of pH on the enzymatic activity was observed. Using the buffers of pH ranging from 4.5 to 9.5, the pH optimum was calculated for the proteolytic activity. The relative activity was calculated at different pHs by taking the activity at pH 7.0 as 100 %. The triplicate data were averaged and normalized using the reaction with no enzyme.

### 5.4.4 Effect of NaCl concentration on enzyme activity

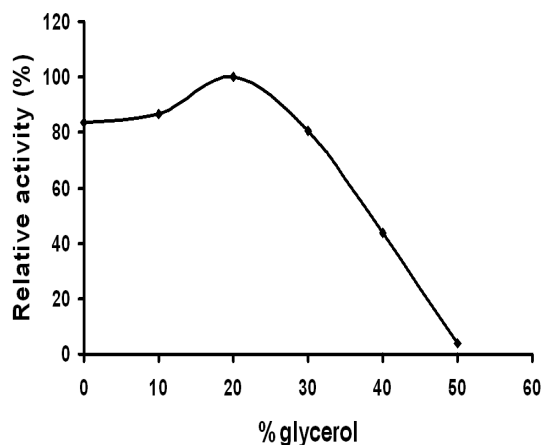
The purified active CVCP was used to study the effect of NaCl concentration on *trans*-proteolytic activity. The estimation of the effect of NaCl concentration on the activity was done by using different NaCl concentrations (25, 50, 100, 200, 300 and 500 mM). The increase in the enzymatic activity was observed from 0 to 0.1 M NaCl concentration while afterwards from 0.1 to 0.5 M a gradual decrease in the activity was observed. The enzyme shows the maximum activity at 100 mM NaCl (Figure 5.4.4.1). The significant increase in enzymatic activity up to 100 mM NaCl concentration might be due to the stabilization of the enzyme with increasing ionic strength, however, the decrease in activity above 100 mM NaCl might be the result of alteration in some interaction at the active site or the specificity pocket at high NaCl concentration.



**Figure 5.4.4.1:** The effect of NaCl concentration on the enzymatic activity was observed. The activity at NaCl concentration of 100 mM was taken as 100 % and the relative activity for other NaCl concentrations was calculated. All the readings are the average of the triplicate data and the values are normalized by using readings obtained from the reaction containing no enzyme.

#### 5.4.5 Glycerol tolerance

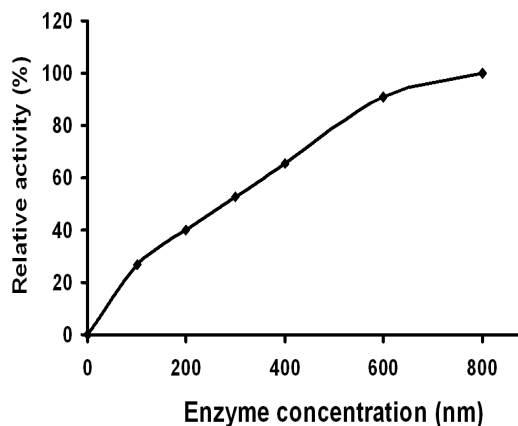
The effect of glycerol on the enzyme activity is also observed and the glycerol has been added to the buffer used in the reaction mixture. Different concentrations of the glycerol (0 to 50 %) were used and the reaction velocity was measured. At 10 % glycerol, negligible increase in the activity is found, however some increase has been noticed at 20 % glycerol concentration. Afterwards increase in the glycerol concentration results in the decrease in activity. At 50 % concentration, the activity becomes almost nil and there is complete loss of enzyme action (Figure 5.4.5.1). The results indicate that up to 30 % glycerol concentration in the reaction buffer can be used, afterwards it inhibits the catalytic activity of enzyme.



**Figure 5.4.5.1:** Dependence of the CVCP proteolytic activity on the glycerol concentration. The effect of increase in glycerol concentration (0 to 50 %) in the reaction buffer on the enzymatic activity of the active CVCP has been observed and the relative activity was measured by taking the activity at 20 % glycerol to be 100 %. The triplicate data were used for the preparation of the graph.

### 5.4.6 Increasing enzyme concentration versus proteolytic activity

To investigate the effect of increasing enzyme concentration on the protease activity, different enzyme concentrations ranging from 0 to 800 nM are used. A linear curve is obtained between increasing enzyme concentration and the relative activity (Figure 5.4.5.2). Thus, with increase in enzyme concentration increased enzyme velocity has been observed.



**Figure 5.4.6.1: Effect of enzyme concentration on the proteolytic activity of the CVCP.** The increasing enzyme concentration results in continuous increase in the enzyme activity. The enzyme concentrations ranging from 0 to 800 nM were used for the reaction. The activity at 800 nM was the maximum, hence this has been considered as 100 % enzyme activity. The enzyme activity at other enzyme concentrations was calculated as the relative activity to the maximum value. The graphs represent the average data obtained from three experiments. The data got from the reaction with no enzyme were used to normalize all the readings.

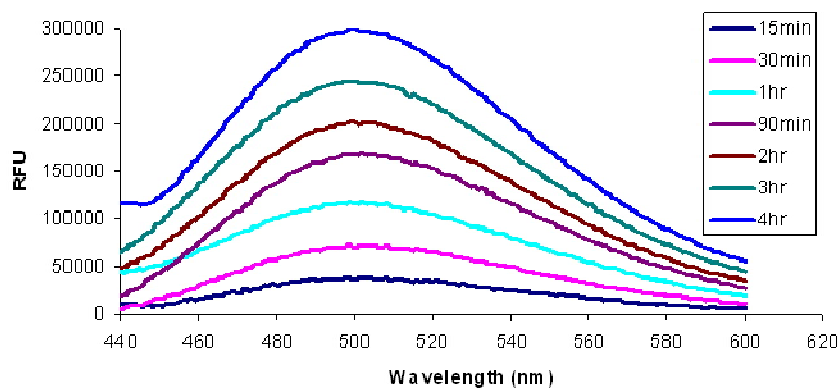
### 5.4.7 Inhibition studies

The tryptophan based inhibitors tryptophan, tryptamine, the common protease inhibitor PMSF (phenylmethanesulfonyl fluoride) and the peptide inhibitors EW and EWS were tested to demonstrate the inhibition of the activity. The enzyme was incubated with the inhibitors for 2 hrs and the assay was performed as described. However, there was no significant increase in the FRET signal in the presence of the inhibitors. The increase in the fluorescence was found to be similar to that of the control reaction with no inhibitor. Thus, it can be concluded that these protease inhibitors do not act as an inhibitor against the capsid protease. Thus, the development of highly specific inhibitors is essential for the inhibition of the capsid proteolytic activity. We did not get inhibition with the use of these inhibitors; however, the peptide EW demonstrates some decrease in the enzyme activity. Though the activity inhibition was not significant enough, this peptide might acts as good inhibitor if used with some modifications. There might be a need to develop more specific inhibitors against CVCP.



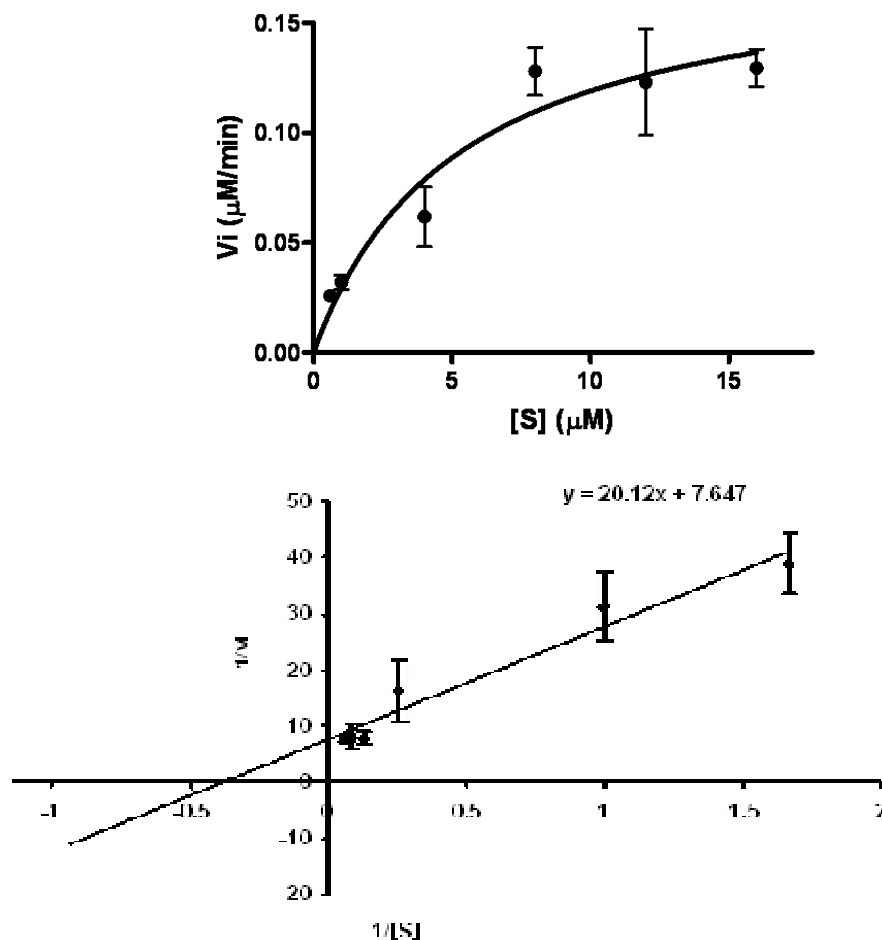
#### 5.4.8 Characterization of *trans*-protease activity for active form of AVCP

For assessment of the *trans*-protease activity of AVCP $\Delta$ 2, the fluorogenic peptide substrate Dabcyl-Gly-Ala-Glu-Glu- **Trp**↓**Ser**-Leu-Ala-Ile-Glu(EDANS) was used. In this fluorometric assay, EDANS is the fluorophore donor having an excitation wavelength at 340 nm and a maximum emission wavelength at 490 nm. DABCYL is the non-fluorescent quencher having a maximum absorption at 470-520 nm wavelength. The activity of AVCP $\Delta$ 2 was measured by FRET analysis at different time points. The donor fluorophore EDANS was excited at 340 nm and emission spectra were recorded by scanning from 440 to 600 nm. To perform this experiment a fixed concentration of substrate (1  $\mu$ M) was used and the fluorescence data measurements were made at specific time intervals. An increase in the intensity of fluorescence with time corresponds to a decrease in the FRET signal, which demonstrates the cleavage of the peptide substrate (Figure 5.4.8.1).



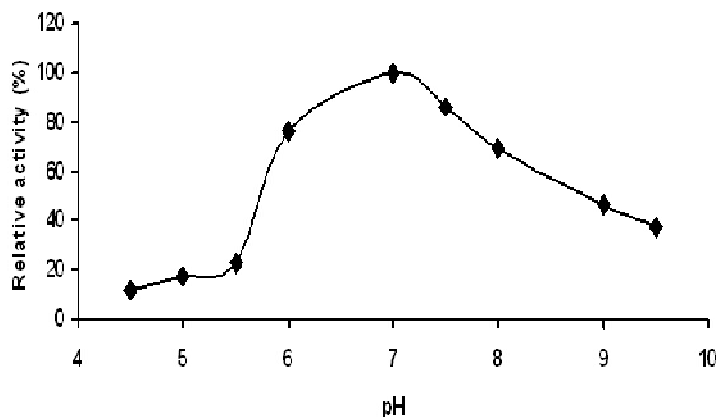
**Figure 5.4.8.1: Analysis of *In vitro trans*-proteolytic activity of AVCP $\Delta$ 2.** The enzymatic cleavage assay was carried out using 15  $\mu$ g of protein in HEPES buffer (20 mM, pH 7.0) by the addition of 1  $\mu$ M fluorogenic peptide substrate at room temperature. The hydrolysis of the peptide substrate was measured at different time intervals. The excitation was done at 340 nm and emission spectrum for each time point is shown as scan from 440 nm to 600 nm.

To determine the kinetic parameters, the calculated  $V_i$  was plotted against different substrate concentrations (Figure 5.4.8.2). A Lineweaver-Burk Plot was used to determine the value of  $K_M$  for the given fluorogenic peptide substrate. The value of  $K_M$  was found to be  $2.63 \pm 0.62$   $\mu$ M and the value for catalytic efficiency  $K_{cat}/K_M$  for the AVCP $\Delta$ 2 *trans*-protease activity was found to be  $4.97 \times 10^4 \text{ M}^{-1} \text{ min}^{-1}$ . The purified inactive form of AVCP containing Trp267 at the C-terminus that blocks the active site, was used as a negative control in all the experiments and showed negligible increase in fluorescence intensity over time.



**Figure 5.4.8.2: Kinetic studies of active AVCP.** The initial velocity ( $\mu\text{M}/\text{min}$ ) was calculated for increasing concentration of the substrate. Different substrate concentrations ranging from  $0.6 \mu\text{M}$  to  $16 \mu\text{M}$  were used (shown on X-axis) and the data were plotted (upper panel). Experiment was done in triplicate and the values represent the average data. All the data were normalized using the same reaction without the enzyme. The lineweaver-burk plot for the activity measurement is shown in lower panel.

The effect of pH on the *trans*-protease activity of AVCP $\Delta$ 2 was determined by performing the fluorometric protease assay in the pH range 4.5 - 9.5. The pH optimal for the peptide substrate (DABCYL)-GAEW $\downarrow$ SLAIE (EDANS) was 7.0 (Figure 5.4.8.3).



**Figure 5.4.8.3:** The influence of the buffer pH on the proteolytic activity of the active AVCP has been observed. The buffers ranging from pH 4.5 to 9.5 were used for the optimization process. The activity was maximum at pH 7.0, hence it was taken as 100 % activity and the activity at other pHs was calculated relative to this value. The data represents the average of the readings taken in triplicate. The data normalization was done using the values of the reaction without enzyme.

## 5.5 Conclusion

A few antiviral drugs have been developed in the last decade (De Clercq *et al.*, 2001), however, viruses are evolving rapidly and more are emerging in different areas. There is a lack of proper treatment against very important viruses. The re-emergence of CHIKV as an epidemic in a number of different regions (Powers & Logue, 2007; Sergon *et al.*, 2007; 2004), the addition in the mosquito vector due to the evolution of virus (mutation in E1 glycoprotein) (Schuffenecker *et al.*, 2006) and the pathogenicity of virus including persistent arthralgia (Larrieu *et al.*, 2010) makes it a significant pathogen which needs proper attention. This gives rise to the necessity of anti-CHIKV drug development and efficient therapeutic protocol manifestation. Though the mortality rate for Chikungunya infection is not so high, the high evolution rate in the genome of CHIKV encourages us for the development of new antivirals and therapeutics against CHIKV.

In last few years, the Chikungunya cases are increasing drastically resulting in ravaging outbreaks. A number of attempts have made to develop the vaccine for the prevention of CHIKV infection, however those can act only before the infection. Once infection occurs, the vaccination cannot help, hence there is an urgent need to develop anti-CHIKV drugs for its treatment. A very less number of studies have been performed till date which aimed for antivirals development against CHIKV. The efforts for the anti-CHIKV drug development results in just the treatment of arthritis to some extent. The non-infectious CHIKV replicon having reporter gene and the cell based phenotypic assay for nsP2 inhibitors of CHIKV have also opened up the possibilities for antiviral screening (Pohjala *et al.*, 2011; Lucas-Hourani *et al.*, 2013). However, the lack of proper

## Chapter 5: trans-Proteolytic activity of Alphavirus capsid protease

*in vitro* assay for antiviral development studies promotes us towards more detailed understanding and development of newer methods for anti-CHIKV compounds screening. In last one year, a number of studies are performed to get the inhibitors for CHIKV replication including Trigocherrierin A and Harringtonine (Kaur *et al.*, 2013; Trigocherrierin, 2014), however no drug has been approved yet. Also, the information is very limited and the denotations of the treatment including dose and duration are not available. The capsid protease has not been investigated yet for anti-CHIKV compound screening.

Our experiments show the presence of monomer for both active and inactive CVCP in solution (Figure 5.4.1.3) as also described for the AVCP (Chapter 2). In 2008, the enzymatic characterization of the esterase activity of the truncated form of SFCP was reported (Morillas *et al.*, 2008). Up until now, no study has reported *in vitro* trans-protease activity of the alphavirus CP. As a consequence, the protease activity of alphavirus CP has not been targeted for antiviral drug discovery. In the current study, we focused on the trans-proteolytic activity assay for CVCP using FRET based fluorogenic peptide substrate. The increase in the velocity occurs with increase in the substrate concentration until it gets saturated. In CVCP, it shows maximum activity at substrate concentration of 10  $\mu\text{M}$  after which it becomes constant. The kinetic parameters for the active CVCP were found as  $K_M = 1.27 \pm 0.34 \mu\text{M}$  and  $K_{cat}/K_M = 5.5 \times 10^4 \text{ M}^{-1} \text{ min}^{-1}$ . The characterization of the protease activity was performed to get the optimum conditions for the proteolytic activity of the active CVCP. In this respect, a number of conditions including NaCl concentration, pH range, glycerol concentration and different enzyme concentrations were used for the measurement of the activity and the relative activity was calculated.

Different physiological conditions (NaCl concentration and pH) depict the optimum enzyme activity at 7.0 pH and 100 mM NaCl concentration. The optimum pH is the neutral pH which corresponds to the pH of mosquito saliva and the human blood. Thus, the optimal activity occurs at the physiological pH of 7.0 to match with the host cell pH. The response of the enzymatic activity with increasing ionic strength indicates the dependence of the protease structure on the salt concentration. The effect of glycerol up to 30 % concentration does not make much difference; however above this concentration, the activity starts decreasing and becomes zero at 50 % of glycerol in the reaction buffer. As anticipated, the increase in the enzyme concentration results in continuous increase in enzyme activity.

In this study, recombinant active AVCP in which two C-terminal P<sub>2</sub> and P<sub>1</sub> residues Glu266 and Trp267 have been removed was used for characterization of the *trans*-protease

activity. The kinetic studies showed  $K_{cat}/K_M$  value of  $4.97 \times 10^4 \text{ M}^{-1} \text{ min}^{-1}$  for the *trans*-protease activity of AVCP $\Delta$ 2. This result indicates that the protease domain of truncated CP is fully active and possesses catalytic properties similar to other chymotrypsin-like serine proteases.

The described *trans*-proteolytic assay can be used further for the high throughput compound screening against CVCP that will lead to the inhibition of first step in structural polyprotein processing. Towards a positive annotation, current investigations will provide a worthwhile route for the development of new anti-CHIKV drugs.



## References

1. Acheson, N. H. & Tamm, I. (1967). Replication of Semliki Forest virus: an electron microscopic study. *Virology* 32, 128–143.
2. 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.
3. Akahata, W., Yang, Z. Y., Andersen, H., Sun, S., Holdaway, H. A., Kong, W. P., Lewis, M. G., Higgs, S., Rossmann, M. G., Rao, S. & Nabel, G. J. (2010). A Virus-Like Particle Vaccine for Epidemic Chikungunya Virus Protects Nonhuman Primates against Infection. *Nat. Med.* 16, 334–338.
4. Aliperti, G. & Schlesinger, M. J. (1978). Evidence for an autoprotease activity of Sindbis virus capsid protein. *Virology* 90, 366–369.
5. 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.
6. Angliker, H., Neumann, U., Molloy, S. S., Thomas, G. (1995) Internally quenched fluorogenic substrate for furin. *Anal. Biochem.* 224, 409–412.
7. Banerjee, M., Khayat, R., Walukiewicz, H. E., Odegard, A. L., Schneemann, A. & Johnson, J. E. (2009). Dissecting the functional domains of a nonenveloped virus membrane penetration peptide. *J. virol.* 83, 6929–6933.
8. Barbato, G., Cicero, D. O., Cordier, F., Narjes, F., Gerlach, B., Sambucini, S., Grzesiek, S., Matassa, V. G., de Francesco, R. & Bazzo, R. (2000). Inhibitor binding induces active site stabilization of the HCV NS3 protein serine protease domain. *EMBO J.* 19, 1195–1206.
9. 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.
10. Barton, D. J., Sawicki, S. G. & Sawicki, D. L. (1991). Solubilization and immunoprecipitation of alphavirus replication complexes. *J. Virol.* 65, 1496–1506.
11. Bourjot, M., Leyssen, P., Neyts, J., Dumontet V. & Litaudon M. (2014). Trigocherrierin A, a Potent Inhibitor of Chikungunya Virus Replication. *Molecules* 19, 3617-3627
12. Brighton, S. W. (1984). Chloroquine Phosphate Treatment of Chronic Chikungunya Arthritis: An Open Pilot Study. *S. Afr. Med. J.* 66, 217–218.
13. Briolant, S., Garin, D., Scaramozzino, N., Jouan, A. & Crance, J. M. (2004). In Vitro Inhibition of Chikungunya and Semliki Forest Viruses Replication by Antiviral Compounds: Synergistic Effect of Interferon-Alpha and Ribavirin Combination. *Antiviral Res.* 61, 111–117.
14. Brown, D.T. (1980). Assembly of alphaviruses. Academic Press, Inc., New York.
15. Carter, P. & Wells, J. A. (1988). Dissection of the catalytic triad of a serine protease. *Nature* 332, 564–568.
16. Chaudhary, N. S., Shee, C., Islam, A., Ahmad, F., Yernool, D., Kumar, P. & Sharma, A. K. (2008). Purification and characterization of a trypsin inhibitor from *Putranjiva roxburghii* seeds. *Phytochemistry*, 69, 2120–2126.
17. Chen, S., Chen, L. L., Luo, H. B., Sun, T., Chen, J., Ye, F., CAI, J. H., Shen, J. K., Shen, X. & Jiang, H. L. (2005). Enzymatic activity characterization of SARS

## References

- coronavirus 3C-like protease by fluorescence resonance energy transfer technique. *Acta Pharmacol. Sin.* 26, 99–106.
18. 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.
  19. 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.
  20. 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.
  21. 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.
  22. Collaborative Computational Project Number 4. (1994). The ccp4 suite: Programs for protein crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* 50, 760–763.
  23. Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* 16, 10881–10890.
  24. Couderc, T., Khandoudi, N., Grandadam, M., Visse, C., Gangneux, N., Bagot, S., Prost, J. F. & Lecuit, M. (2009). Prophylaxis and Therapy for Chikungunya Virus Infection. *J. Infect. Dis.* 200, 516–523.
  25. Cross, J. B., Thompson, D. C., Rai, B. K., Baber, J. C., Fan, K. Y., Hu, Y. & Humblet, C. (2009). Comparison of several molecular docking programs: pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.* 49, 1455–1474.
  26. Cruz, D. J. M., Bonotto, R. M., Gomes, R. G., da Silva, C. T., Taniguchi, J. B., No, J. H., Lombardot, B., Schwartz, O., Hansen, M. A. E. & Freitas-Junior, L. H. (2013). Identification of Novel Compounds Inhibiting Chikungunya Virus-Induced Cell Death by High Throughput Screening of a Kinase Inhibitor Library. *PLoS Negl. Trop. Dis.* 7, e2471.
  27. Cui, J., Marankan, F., Fu, W., Crich, D., Mesecar, A. & Johnson, M. E. (2002). An oxyanion-hole selective serine protease inhibitor in complex with trypsin. *Bioorg. & Med. Chem. Lett.* 10, 41–46.
  28. Dafforn, T. R. & Smith, C. J. (2004). Natively unfolded domains in endocytosis: hooks, lines and linkers. *EMBO Rep.* 5, 1046–1052.
  29. Dave, S., Kaur, N. J., Nanduri, R., Dkhar, H. K., Kumar, A., & Gupta, P. (2012). Inhibition of adipogenesis and induction of apoptosis and lipolysis by stem bromelain in 3T3-L1 adipocytes. *PloS one* 7, e30831.
  30. Dave, S., Mahajan, S., Chandra, V., & Gupta, P. (2011). Trifluoroethanol stabilizes the molten globule state and induces non-amyloidic turbidity in stem bromelain near its isoelectric point. *Int. J. Boil. Macromol.* 49, 536–542.
  31. De Clercq, E., Naesens, L., De Bolle, L., Schols, D., Zhang, Y. & Neyts, J. (2001). Antiviral agents active against human herpesviruses HHV-6, HHV-7 and HHV-8. *Rev. Med. Virol.* 11, 381–395.
  32. de Groot, R. J., Rumenapf, T., Kuhn, R. J., Strauss, E. G. & Strauss, J. H. (1991). Sindbis virus RNA polymerase is degraded by the N-end rule pathway. *Proc. Natl. Acad. Sci. USA* 88, 8967–8971.
  33. De Lamballerie, X., Boisson, V., Reynier, J. C., Enault, S., Charrel, R. N., Flahault, A., Roques, P. & Le Grand, R. (2008). On Chikungunya Acute Infection and Chloroquine Treatment. *Vector Borne Zoonotic Dis.* 8, 837–839.



34. Delano, W. L. (2002). The PyMOL molecular graphics system (DeLano Scientific, LLC, Palo Alto, CA).
35. DeTulleo, L. & Kirchhausen, T. (1998). The clathrin endocytic pathway in viral infection. *EMBO J.* 17, 4585–4593.
36. di Marco, S., Rizzi, M., Volpari, C., Walsh, M. A., Narjes, F., Colarusso, S., de Francesco, R., Matassa, V. G. & Sollazzo, M. J. (2000). Inhibition of the hepatitis C virus NS3/4A protease. The crystal structures of two protease-inhibitor complexes. *J. Biol. Chem.* 275, 7152–7157.
37. Dyson, H. J. & Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nature Rev. Mol. Cell. Biol.* 6, 197–208.
38. Egesten, A., Olin, A. I., Linge, H. M., Yadav, M., Mörgelin, M., Karlsson, A., & Collin, M. (2009). SpeB of *Streptococcus pyogenes* differentially modulates antibacterial and receptor activating properties of human chemokines. *PLoS One*, 4, e4769.
39. Eguchi, M. & Kuriyama, K. (1985). Purification and characterization of membrane-bound alkaline proteases from midgut tissue of the silkworm, *Bombyx mori*. *J. Biochem. (Tokyo)* 97, 1437–1445.
40. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2126–2132.
41. Farnum, M. & Zukoski, C. (1999). Effect of Glycerol on the Interactions and Solubility of Bovine Pancreatic Trypsin Inhibitor. *Biophys. J.* 76, 2716–2726.
42. Fattori, D., Urbani, A., Brunetti, M., Ingenito, R., Pessi, A., Prendergast, K., Narjes, F., Matassa, V. G., de Francesco, R. & Steinkuhler, C. J. (2000). Probing the active site of the hepatitis C virus serine protease by fluorescence resonance energy transfer. *J. Biol. Chem.* 275, 15106–15113.
43. Ferlenghi, I., Gowen, B., de Haas, F., Mancini, E. J., Garoff, H., Sjöberg, M. & Fuller, S. D. (1998). The first step: activation of the Semliki Forest virus spike protein precursor causes a localized conformational change in the trimeric spike. *J. Mol. Biol.* 283, 71–81.
44. Fink, A. L. (2005). Natively unfolded proteins. *Curr. Opin. Struct. Biol.* 15, 35–41.
45. Forsell, K., Xing, L., Kozlovskaya, T., Cheng, R. H. & Garoff, H. (2000). Membrane proteins organize a symmetrical virus. *EMBO J.* 19, 5081–5091.
46. Forsell, K., Griffiths, G. & Garoff, H. (1996). Preformed cytoplasmic nucleocapsids are not necessary for alphavirus budding. *EMBO J.* 15, 6495–6505.
47. Forsell, K., Suomalainen, M. & Garoff, H. (1995). Structure-function relation of the NH<sub>2</sub>-terminal domain of the Semliki Forest virus capsid protein. *J. Virol.* 69, 1556–1563.
48. Friedman, P. N., Levin, J. G., Grimley, P. M. & Brezesky, I. K. (1972). Membrane-associated replication complex in arbovirus infection. *J. Virol.* 10, 504–515.
49. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A. & Klicic, J. J. (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* 47, 1739–1749.
50. 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.
51. Frolov, I., & Schlesinger, S. (1994). Translation of Sindbis virus mRNA: Effects of sequences downstream of the initiating codon. *J. Virol.* 68, 8111–8117.

## References

52. Frolova, E., Frolov, I. & Schlesinger, S. (1997). Packaging signals in alphaviruses. *J. Virol.* 71, 248–258.
53. Froshauer, S., Kartenbeck, J. & Helenius, A. (1988). Alphavirus RNA replicase is located on the cytoplasmic surface of endosomes and lysosomes. *J. Cell Biol.* 107, 2075–2086.
54. 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.
55. 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.
56. 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.
57. 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.
58. Gardner, J., Anraku, I., Le, T. T., Larcher, T., Major, L., Roques, P., Schroder, W. A., Higgs, S. & Suhrbier, A. (2010). Chikungunya Virus Arthritis in Adult Wild-Type Mice. *J. Virol.* 84, 8021–8032.
59. 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.
60. 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.
61. 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.
62. Garoff, H., & Simons, K. (1974). Location of the spike glycoproteins in the Semliki Forest virus membrane. *Proc. Natl. Acad. Sci. USA* 71, 3988–3992.
63. 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.
64. Gekko, K. & Timasheff, S. N. (1981). Thermodynamic and Kinetic Examination of Protein Stabilization by Glycerol. *Biochemistry* 20, 4677–4686.
65. Gekko, K. & Timasheff, S. N. (1981). Mechanism of Protein Stabilization by Glycerol: Preferential Hydration in Glycerol-Water Mixtures. *Biochemistry* 20, 4667–4676.
66. Glanville, N. & Ulmanen, J. (1976). Biological activity of in vitro synthesized protein: binding of Semliki Forest virus capsid protein to the large ribosomal subunit. *Biochem. Biophys. Res. Commun.* 71, 393–399.
67. 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.
68. Gonzalez, M. E. & Carrasco, L. (2003). Viroporins. *FEBS Lett.* 552, 28–34.
69. Gouet, P., Courcelle, E., Stuart, D. I. & Metz, F. (1999). ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* 15, 305–308.

70. Griffiths, G., Fuller, S. D., Back, R., Hollinshead, M., Pfeiffer, S. & Simons, K. (1989). The dynamic nature of the Golgi complex. *J. Cell Biol.* 108, 277–297.
71. Griffiths, G., Quinn, P. & Warren, G. (1983). Dissection of the Golgi complex. I. Monensin inhibits the transport of viral membrane proteins from medial to trans Golgi cisternae in baby hamster kidney cells infected with Semliki Forest virus. *J. Cell Biol.* 96, 835–850.
72. Grimley, P. M., Levin, J. G., Berezsky, I. K. & Friedman, R. M. (1972). Specific membranous structures associated with the replication of group A arboviruses. *J. Virol.* 10, 492–503.
73. Grimley, P. M., Berezsky, I. K. & Friedman, P. N. (1968). Cytoplasmic structures associated with an arbovirus infection: loci of viral ribonucleic acid synthesis. *J. Virol.* 2, 1326–1338.
74. 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.
75. 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.
76. 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.
77. Hahn, C. S., Lustig, S., Strauss, E. G. & Strauss, J. H. (1988). Western equine encephalitis virus is a recombinant virus. *Proc. Natl Acad. Sci. USA* 85, 5997–6001.
78. 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.
79. 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.
80. 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.
81. Hedstrom, L. (2002). Serine protease mechanism and specificity. *Chem. rev.* 102, 4501–4524.
82. 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.
83. Hernandez, R., Luo, T. & Brown, D. T. (2001). Exposure to low pH is not required for penetration of mosquito cells by Sindbis virus. *J. Virol.* 75, 2010–2013.
84. Hoarau, J. J., Jaffar Bandjee, M. C., Trotot, P. K., Das, T., Li-Pat-Yuen, G., *et al.* (2010). Persistent chronic inflammation and infection by Chikungunya arthritogenic alphavirus in spite of a robust host immune response. *J. Immunol.* 184, 5914–5927.
85. Holm, L. & Rosenström, P. (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Res.* 38, 545–549.
86. 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.
87. Inoue, H., Nojima, H. & Okayama, H. (1990). High efficiency transformation of *Escherichia coli* with plasmids. *Gene* 96, 23–28.
88. Ivanova, L., Lustig, S. & Schlesinger, M. J. (1999). A pseudo-revertant of a Sindbis virus 6K protein mutant, which corrects for aberrant particle formation, contains two

## References

- new mutations that map to the ectodomain of the E2 glycoprotein. *Virology* 206, 1027–1034.
89. 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.
90. 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.
91. Jin, S., Ellis, E., Veetil, J. V., Yao, H. & Ye, K. (2011). Visualization of human immunodeficiency virus protease inhibition using a novel Forster resonance energy transfer molecular probe. *Biotechnol. Prog.* 27, 1107–1114.
92. Johnson, G. & Moore, S. W. (2002). Catalytic anti-bodies with acetylcholinesterase activity. *J. Immunol. Methods* 269, 13–28.
93. Johnson, G. & Moore, S.W. (2000). Cholinesterase-like catalytic antibodies: reaction with substrates and inhibitors. *Mol. Immunol.* 37, 707–719.
94. Johnson, J. E. & Banerjee, M. (2008). Activation, exposure and penetration of virally encoded, membrane-active polypeptides during non-enveloped virus entry. *Curr. Protein Pep. Sci.* 9, 16–27.
95. Jones, K. J., Scupham, R. K., Pfeil, J. A., Wan, K., Sagik, B. P. & Bose, H. R. (1977). Interaction of Sindbis virus glycoproteins during morphogenesis. *J. Virol.* 21, 778–787.
96. 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.
97. Jung, G., Ueno, H., Hayashi, R. & Liao, T. H. (1995). Identification of the catalytic histidine residue participating in the charge-relay system of carboxypeptidase Y. *Protein Sci.* 4, 2433–2435.
98. 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.
99. 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.
100. Kaur, P., Thiruchelvan, M., Lee, R. C., Chen, H., Chen, K. C., Ng, M. L. & Chu, J. J. (2013). Inhibition of chikungunya virus replication by harringtonine, a novel antiviral that suppresses viral protein expression. *Antimicrob. Agents Chemother.* 57, 155–167.
101. Kielian, M. (2002). Structural surprises from the flaviviruses and alphaviruses. *Mol. Cell* 9, 454–456.
102. Kielian, M. (1995). Membrane fusion and the alphavirus life cycle. *Adv. Virus Res.* 45, 113–151.
103. Kielian, M. & Helenius, A. (1985). pH-Induced alterations in the fusogenic spike protein of Semliki Forest virus. *J. Cell Biol.* 101, 2284–2291.
104. Kim, H. Y., Kuhn, R. J., Patkar, C., Warriar, R. & Cushmana, 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.
105. Kim, H. Y., Patkar, C., Warriar, R., Kuhn, R. J. & Cushmana, 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.

106. Kim, J. L., Morgenstern, K. A., Lin, C., Fox, T., Dwyer, M. D., Landro, J. A. *et al.* (1996). Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide. *Cell* 87, 343–355.
107. 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.
108. 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.
109. 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.
110. Knight, C. G., Willenbrock, F. & Murphy, G. (1992). A novel coumarin-labelled peptide for sensitive continuous assays of the matrix metalloproteinases. *FEBS Lett.* 296, 263–6.
111. 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.
112. Krem, M. M., Prasad, S. & Cera, E. D. (2002). Ser214 Is Crucial for Substrate Binding to Serine Proteases. *J. Biol. Chem.* 277, 40260–40264.
113. Krissinel, E. & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372, 774–797.
114. Kuhn, R.J. (2007). Togaviridae: the viruses and their replication. In: Knipe, D.M., Howley, P.M. (Eds.), *Fields' Virology*, fifth ed. Lippincott, Williams and Wilkins, New York, 1001–1022.
115. Kuhn, R. J., Hong, Z. & Strauss, J. H. (1990). Mutagenesis of the 3' nontranslated region of Sindbis virus RNA. *J. Virol.* 64, 1465–1476.
116. Kujala, P., Ikaheimonen, A., Ehsani, N., Vihinen, H., Auvinen, P. & Kaariainen, L. (2001). Biogenesis of the Semliki Forest virus RNA replication complex. *J. Virol.* 75, 3873–3884.
117. Kumar, P., Singh, M. & Karthikeyan, S. (2011). Crystal structure analysis of icosahedral lumazine synthase from *Salmonella typhimurium*, an antibacterial drug target. *Acta Crystallogr. D Biol. Crystallogr.* 67, 131–139.
118. Kumar, P., Singh, M., Gautam, R. & Karthikeyan, S. (2010). Potential anti-bacterial drug target: Structural characterization of 3, 4-dihydroxy-2-butanone-4-phosphate synthase from *Salmonella typhimurium* LT2. *Proteins* 78, 3292–3303.
119. Kumar, A., Sharma, J., Grover, S., Kumar Mohanty, A. & Batish, V. K. (2007). Molecular cloning and expression of goat (*Capra hircus*) prochymosin in *E. coli*. *Food Biotechnol.* 21, 57–69.
120. 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.
121. Landro, J. A., Raybuck, S. A., Luong, Y. P., O'Malley, E. T., Harbeson, S. L., Morgenstern, K. A., Govinda, R. & Livingston, D. J. (1997). Mechanistic role of an NS4A peptide cofactor with the truncated NS3 protease of hepatitis C virus: elucidation of the NS4A stimulatory effect via kinetic analysis and inhibitor mapping. *Biochemistry* 36, 9340–9348.

## References

122. 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.
123. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 26, 283–291.
124. Lee, S., Kuhn, R. J. & Rossmann, M. G. (1998). Probing the potential glycoprotein binding site of sindbis virus capsid protein with dioxane and model building. *Proteins* 33, 311–317.
125. 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.
126. 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.
127. Lee, H., Ricker, P. D. & Brown, D. T. (1994). The configuration of Sindbis virus envelope proteins is stabilized by the nucleocapsid protein. *Virology* 204, 471–474.
128. Lenglet, Y., Barau, G., Robillard, P. Y., Randrianaivo, H., Michault, A., Bouveret, A., Gérardin, P. *et al.* (2006). Chikungunya infection in pregnancy: evidence for intrauterine infection in pregnant women and vertical transmission in the parturient, Survey of the Reunion Island outbreak. *J. Gynecol. Obstet. Biol. Reprod. (Paris)* 35, 578–583.
129. Lescar, J., Roussel, A., Wien, M. W., Navaza, J., Fuller, S. D. & Wengler, G. (2001). The Fusion glycoprotein shell of Semliki Forest virus: an icosahedral assembly primed for fusogenic activation at endosomal pH. *Cell* 105, 137–148.
130. Leslie, A. G. W. & Powell, H. R. (2007). Processing diffraction data with Mosflm. *Evolv. Methods Macromol. Crystallograph.* 245, 41–51.
131. 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.
132. 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.
133. Li, Y. & Jing, G. (2000). Double point mutant F34W/W140F of staphylococcal nuclease is in a molten globule state but highly competent to fold into a functional conformation. *J. Biochem. (Tokyo)* 128, 739–744.
134. 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.
135. 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.
136. Liang, Y., Ye, H., Kang, C. B. & Yoon, H. S. (2007). Domain 2 of nonstructural protein 5A (NS5A) of hepatitis C virus is natively unfolded. *Biochemistry* 46, 11550–11558.
137. Ligon, B. L. (2006). Reemergence of an unusual disease: the chikungunya epidemic. *Semin. Pediatr. Infect. Dis.* 17, 99–104.

138. 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.
139. 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.
140. 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.
141. Lin, C. W., Tsai, C. H., Tsai, F. J., Chen, P. L., Lai, C. C., Wan, L., Chiu, H. H. & Lin, K. H. (2004). Characterization of trans- and cis-cleavage activity of the SARS coronavirus 3CLpro protease: basis for the in vitro screening of anti-SARS drugs. *FEBS Lett.* 574, 131–137.
142. Liu, N. & Brown, D. T. (1993). Phosphorylation and dephosphorylation events play critical roles in Sindbis virus maturation. *Virology* 196, 703–711.
143. 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.
144. 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.
145. 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.
146. 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.
147. Lottenberg, R., Christensen, U., Jackson, C. M. & Coleman, P. L. (1981). Assay of coagulation proteases using peptide chromogenic and fluorogenic substrates. *Methods Enzymol.* 80, 341–361.
148. 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.
149. Lusa, S., Garoff, H. & Liljestrom, P. (1991). Fate of the 6K membrane protein of Semliki Forest virus during virus assembly. *Virology* 185, 843–846.
150. Luthy, R., Bowie, J. U. & Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. *Nature* 356, 83–85.
151. Madan, V., Castello, A. & Carrasco, L. (2008). Viroporins from RNA viruses induce caspase-dependent apoptosis. *Cell. Microbiol.* 10, 437–451.
152. Madan, V., Sanz, M. A. & Carrasco, L. (2005). Requirement of the vesicular system for membrane permeabilization by Sindbis virus. *Virology* 332, 307–315.
153. Mahajan, S., Chandra, V., Dave, S., Nanduri, R. & Gupta, P. (2012). Stem Bromelain-Induced Macrophage Apoptosis and Activation Curtail Mycobacterium tuberculosis Persistence. *J. Infect. Dis.* 206, 366–376.
154. 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.

## References

155. 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.
156. Marsh, M., Bolzau, E. & Helenius, A. (1983). Penetration of Semliki Forest virus from acidic prelysosomal vacuoles. *Cell* 32, 931–940.
157. 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.
158. Matthews, D. A., Dragovich, P. S., Webber, S. E., Fuhrman, S. A., Patick, A. K., Zalman, L. S., Hendrickson, T. F., *et al.* (1999). Structure-assisted design of mechanism-based irreversible inhibitors of human rhinovirus 3C protease with potent antiviral activity against multiple rhinovirus serotypes. *Proc. Natl. Acad. Sci. USA* 96, 11000–11007.
159. McKinlay, M. A. (2001). Recent advances in the treatment of rhinovirus infections. *Curr. Opin. Pharmacol.* 1, 477–481.
160. 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.
161. Melancon, P. & Garoff, H. (1987). Processing of the Semliki Forest virus structural polyprotein: role of the capsid protease. *J. Virol.* 61, 1301–1309.
162. Mellman, I., Fuchs, R. & Helenius, A. (1986). Acidification of the endocytic and exocytic pathways. *Annu. Rev. Biochem.* 55, 663–700.
163. Metallo, S. J. (2010). Intrinsically disordered proteins are potential drug targets. *Curr. Opin. Chem. Biol.* 14, 481–488.
164. 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.
165. Mi, S. & Stollar, V. (1991). Expression of Sindbis virus nsP1 and methyltransferase activity in *Escherichia coli*. *Virology* 184, 423–427.
166. 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.
167. 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.
168. 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
169. 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.
170. Mukhopadhyay, S., Zhang, W., Gabler, S., *et al.* (2006). Mapping the structure and function of the E1 and E2 glycoproteins in alphaviruses. *Structure* 14, 63–73.



171. 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.
172. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogr. D Biol. Crystallogr.* 53, 240–255.
173. 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.
174. Narayanan, A., Kumar, S., Evrard, A. N., Paul, L. N. & Yernool, D. A. (2014). An asymmetric heterodomain interface stabilizes a response regulator–DNA complex. *Nat. Commun.* 5.
175. Narayanan, A., Paul, L. N., Tomar, S., Patil, D. N., Kumar, P. & Yernool, D. A. (2012). Structure-function studies of DNA binding domain of response regulator KdpE reveals equal affinity interactions at DNA half-sites. *PloS one* 7, e30102.
176. Navaratnarajah, C. K. & Kuhn, R. J. (2007). Functional characterization of the Sindbis virus E2 glycoprotein by transposon linker-insertion mutagenesis. *Virology* 363, 134–147.
177. 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.
178. Nitsche, C. & Klein, C. D. (2013). Fluorimetric and HPLC-Based Dengue Virus Protease Assays Using a FRET Substrate. In *Antiviral Methods and Protocols* 221–236.
179. Odake, S., Kam, C. M., Narasimhan, L., Poe, M., Blake, J. T., Krahenbuhl, O., Tschopp, J. & Powers, J. C. (1991). Human and murine cytotoxic T lymphocyte serine proteases: subsite mapping with peptide thioester substrates and inhibition of enzyme activity and cytolysis by isocoumarins. *Biochemistry* 30, 2217–2227.
180. 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.
181. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol.* 276, 307–326.
182. 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.
183. 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.
184. 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.
185. Ozden, S., Lucas-Hourani, M., Ceccaldi, P. E., Basak, A., Valentine, M., Benjannet, S., Hamelin, J., *et al.* (2008). Inhibition of Chikungunya Virus Infection in Cultured Human Muscle Cells by Furin Inhibitors: Impairment of the Maturation of the E2 Surface Glycoprotein. *J. Biol. Chem.* 283, 21899–21908.
186. Padbidri, V. S. & Gnaneswar, T. T. (1979). Epidemiological investigations of chikungunya epidemic at Barsi, Maharashtra state, India. *J. Hyg. Epidemiol. Microbiol. Immunol.* 23, 445–451.

## References

187. 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.
188. 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.
189. 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.
190. 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.
191. 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.
192. Patick, A. K. & Potts, K. E. (1998). Protease Inhibitors as Antiviral Agents. *Clin. Microbiol. Rev.* 11, 614–627.
193. Patil, D. N., Chaudhary, A., Sharma, A. K., Tomar, S. & Kumar, P. (2012). Structural basis for dual inhibitory role of tamarind Kunitz inhibitor (TKI) against factor Xa and trypsin. *FEBS J.* 279, 4547–4564.
194. 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.
195. 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.
196. 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.
197. 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. *J. Comput. Chem.* 25, 1605–1612.
198. 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.
199. Pohjala, L., Utt, A., Varjak, M., Lulla, A., Merits, A., Ahola, T. & Tammela, P. (2011). Inhibitors of alphavirus entry and replication identified with a stable Chikungunya replicon cell line and virus-based assays. *PLoS One* 6, e28923.
200. Powers, A. M. & Logue, C. H. (2007). Changing patterns of chikungunya virus: re-emergence of a zoonotic arbovirus. *J. Gen. Virol.* 88, 2363–2377.
201. Powers, J. C. & Kam, C. M. (1995). Peptide thioester substrates for serine peptidases and metalloendopeptidases. *Methods Enzymol.* 248, 3–18.
202. 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.
203. 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.

204. Ravichandran, R. & Manian, M. (2008). Ribavirin Therapy for Chikungunya Arthritis. *J. Infect. Dev. Ctries.* 2, 140–142.
205. Reichert, E., Clase, A., Bacetty, A. & Larsen, J. (2009). Alphavirus Antiviral Drug Development: Scientific Gap Analysis and Prospective Research Areas. *Biosecur Bioterror.* 7, 413–427.
206. 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.
207. Rice, C. M. & Strauss, J. H. (1982). Association of Sindbis virion glycoproteins and their precursors. *J. Mol. Biol.* 154, 325–348.
208. 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.
209. Rice, C. M. & Strauss, J. H. (1981). Synthesis, cleavage, and sequence analysis of DNA complementary to the 26S messenger RNA of Sindbis virus. *J. Mol. Biol.* 150, 315–340.
210. Rikkinen, M., Peränen, J. & Kääriäinen, L. (1994). ATPase and GTPase activities associated with Semliki Forest virus non-structural protein nsP2. *J. Virol.* 68, 5804–5810.
211. Robin, S., Ramful, D., Le Seach, F., Jaffar-Bandjee, M. C., Rigou, G. & Alessandri, J. L. (2008). Neurologic manifestations of pediatric Chikungunya infection. *J. Child. Neurol.* 23, 1028–1035.
212. Ross, R. W. (1956). The Newala epidemic. III. The virus: isolation, pathogenic properties and relationship to the epidemic. *J. Hyg. (Lond)* 54, 177–191.
213. Rossmann, M. G. & Johnson, J. E. (1989). Icosahedral RNA virus structure. *Ann. Rev. Biochem.* 58, 533–573.
214. Rubach, J. K., Wasik, B. R., Rupp, J. C., Kuhn, R. J., Hardy, R. W., & Smith, J. L. (2009). Characterization of purified Sindbis virus nsP4 RNA-dependent RNA polymerase activity *in vitro*. *Virology* 384, 201–208.
215. Rümenapf, T., Strauss, E. G. & Strauss, J. H. (1995). Aura virus is a New World representative of Sindbis-like viruses. *Virology* 208, 621–633.
216. 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.
217. Rümenapf, T., Strauss, E. G. & Strauss, J. H. (1994). Subgenomic mRNA of Aura alphavirus is packaged into virions. *J. Virol.* 68, 56–62.
218. Rumthao, S., Lee, O., Sheng, Q., Fu, W. T., Mulhearn, D. C., Crich, D., Mesecar, A. D. & Johnson, M. E. (2004). Design, synthesis, and evaluation of oxyanion-hole selective inhibitor substituents for the S1 subsite of factor Xa. *Bioorg. & Med. Chem. Lett.* 14, 5165–5170.
219. 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.
220. Ryan, C., Ivanova, L. & Schlesinger, M. J. (1998). Mutations in the Sindbis virus capsid gene can partially suppress mutations in the cytoplasmic domain of the virus E2 glycoprotein spike. *Virology* 243, 380–387.
221. 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

## References

- neurovirulence of neuroadapted and nonneuroadapted Sindbis viruses. *J. Virol.* 81, 3563–3573.
222. 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.
223. Sabariego, R., Picazo, F., Domingo, B., Franco, S., Martinez, M.A. & Llopis, J. (2009). Fluorescence resonance energy transfer-based assay for characterization of hepatitis C virus NS3-4A protease activity in live cells. *Antimicrob. Agents Chemother.* 53, 728–734.
224. Sali, A. & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779–815.
225. 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.
226. 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.
227. 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.
228. Schuffenecker, I., Itean, I., Michault, A., Murri, S., Frangeul, L., Vaney, M. C. et al. (2006). Genome microevolution of chikungunya viruses causing the Indian Ocean outbreak. *PLoS Med.* 3, e263
229. Schuttelkopf, A. W. & van Aalten, D. M. (2004). PRODRG: A tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. D Biol. Crystallogr.* 60, 1355–1363.
230. Schwartz, O. & Matthew, L. A. (2010). Biology and pathogenesis of chikungunya virus. *Nat. Rev. Microbiol.* 8, 491–500.
231. Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. (2003). SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res.* 31, 3381–3385.
232. Sergon, K., Njuguna, C., Kalani, R., Ofula, V., Onyango, C., Konongoi, L. S. et al. (2008). Seroprevalence of chikungunya virus (CHIKV) infection on Lamu Island, Kenya, October 2004. *Am. J. Trop. Med. Hyg.* 78, 333–337.
233. Sergon, K., Yahaya, A. A., Brown, J., Bedja, S. A., Mlindasse, M., Agata, N. et al. (2007). Seroprevalence of Chikungunya virus infection on Grande Comore Island, union of the Comoros, 2005. *Am. J. Trop. Med. Hyg.* 76, 1189–1193.
234. 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.
235. Sherman, L. A. & Griffin, D. E. (1990). Pathogenesis of encephalitis induced in newborn mice by virulent and avirulent strains of Sindbis virus. *J. Virol.* 64, 2041–2046.
236. 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.
237. 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.
238. Sidorowicz, W., Jackson, G. C. & Behal, F. J. (1980). Multiple molecular forms of human pancreas alanine aminopeptidase. *Clin. Chim. Acta* 104, 169–179.
  239. Simon, F., Javelle, E., Oliver, M., Leparç-Goffart, I. & Marimoutou, C. (2011). Chikungunya virus infection. *Curr. Infect. Dis. Rep.* 13, 218–228.
  240. Singh, M., Kumar, P. & Karthikeyan, S. (2011). Structural basis for pH dependent monomer–dimer transition of 3, 4-dihydroxy 2-butanone-4-phosphate synthase domain from *Mycobacterium tuberculosis*. *J. Struct. Boil.* 174, 374–384.
  241. Singh, I. & Helenius, A. (1992). Role of ribosomes in Semliki Forest virus nucleocapsid uncoating. *J. Virol.* 66, 7049–7058.
  242. 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.
  243. 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.
  244. 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.
  245. Skoging, U., Vihinen, M., Nilsson, L. & Liljestrom, P. (1996). Aromatic interactions define the binding of the alphavirus spike to its nucleocapsid. *Structure* 4, 519–529.
  246. Snyder, J. E., Kulcsar, K. A., Schultz, K. L., Riley, C. P., Neary, J. T., Marr, S., Jose, J., Griffin, D. E. & Kuhn, R. J. (2013). Functional characterization of the alphavirus TF protein. *J. Virol.* 87, 8511–8523.
  247. Söderlund, H. & Ulmanen, I. (1977). Transient association of Semliki Forest virus capsid protein with ribosomes. *J. Virol.* 24, 907–909.
  248. Söderlund, H. (1973). Kinetics of formation of Semliki Forest virus nucleocapsid. *Intervirology* 1, 354–361.
  249. 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.
  250. Sousa, R. (1995). Use of glycerol, polyols and other protein structure stabilizing agents in protein crystallization. *Acta Crystallogr. D Biol. Crystallogr.* 51, 271–277.
  251. Strauss, J. H. & Strauss, E. G. (2001). Virus evolution: how does an enveloped virus make a regular structure? *Cell* 105, 5–8.
  252. Strauss, J. H., Strauss, E. G. & Kuhn, R. J. (1995). Budding of alphaviruses. *Trends Microbiol.* 3, 346–350.
  253. Strauss, J. H. & Strauss, E. G. (1994). The alphaviruses: gene expression, replication, and evolution. *Microbiol. Rev.* 58, 491–562.
  254. Strauss, E. G., Rice, C. M. & Strauss, J. H. (1984). Complete nucleotide sequence of the genomic RNA of Sindbis virus. *Virology* 133, 92–110.
  255. 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.
  256. Suomalainen, M. & Garoff, H. (1992). Alphavirus spike-nucleocapsid interaction and network antibodies. *J. Virol.* 66, 5106–5109.

## References

257. Suomalainen, M., Liljestrom, P. & Garoff, H. (1992). Spike protein-nucleocapsid interactions drive the budding of alphaviruses. *J. Virol.* 66, 4737–4747.
258. 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.
259. 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.
260. 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. *Virol. J.* 7, 327.
261. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680.
262. 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.
263. 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.
264. Tucker, P. C., Lee, S. H., Bui, N., Martinie, D. & Griffin, D. E. (1997). Amino acid changes in the Sindbis virus E2 glycoprotein that increase neurovirulence improve entry into neuroblastoma cells. *J. Virol.* 71, 6106–6112.
265. Uchime, O., Fields, W. & Kielian, M. (2013). The Role of E3 in pH Protection during Alphavirus Assembly and Exit. *J. Virol.* 87, 10255–10262.
266. Ulmanen, I., Soderlund, H. & Kääriäinen, L. (1976). Semliki Forest virus capsid protein associates with the 60S ribosomal subunit in infected cells. *J. Virol.* 20, 203–210.
267. Uversky, V. N., Oldfield, C. J. & Dunker, A. K. (2008). *Annu. Rev. Biophys.* 37, 215–246.
268. Uversky, V. N. (2002). Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* 11, 739–756.
269. 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.
270. Vagin, A. & Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* 30, 1022–1025.
271. 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.
272. 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.
273. Vaux, D. J. T., Helenius, A. & Mellman, I. (1988). Spike-nucleocapsid interaction in Semliki Forest virus reconstructed using network antibodies. *Nature* 336, 36–42.

274. 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.
275. von Bonsdorff, C. H. & Harrison, S. C. (1978). Hexagonal glycoprotein arrays from Sindbis virus membranes. *J. Virol.* 28, 578–583.
276. von Bonsdorff, C. H. & Harrison, S. C. (1975). Sindbis virus glycoproteins form a regular icosahedral surface lattice. *J. Virol.* 16, 141–145.
277. 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.
278. 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.
279. 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.
280. Wang, Q. M. (1999). Protease inhibitors as potential antiviral agents for the treatment of picornaviral infections. *Prog. Drug Res.* 52, 197–219.
281. 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.
282. 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.
283. 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.
284. 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.
285. 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.
286. Wengler, G. & Wengler, G. (1984). Identification of a transfer of viral core protein to cellular ribosomes during the early stages of alphavirus infection. *Virology* 134, 435–440.
287. 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.
288. White, J. & Helenius, A. (1980). pH-dependent fusion between the Semliki Forest virus membrane and liposomes. *Proc. Natl. Acad. Sci. U.S.A.* 77, 3273–3277.
289. 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.
290. 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.

## References

291. Winn, M., Isupov, M. & Murshudov, G. N. (2000). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D Biol. Crystallogr.* 57, 122–133.
292. 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.
293. 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.
294. 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.
295. 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.
296. 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.
297. 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.
298. 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.
299. 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.
300. 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.
301. Zhao, H. & Garoff, H. (1992). Role of cell surface spikes in alphavirus budding. *J. Virol.* 66, 7089–7095.
302. Ziemiecki, A., Garoff, H. & Simons, K. (1980). Formation of the Semliki Forest virus membrane glycoprotein complexes in the infected cell. *J. Gen. Virol.* 50, 111–123.